

PRACTICE QUESTIONS

Oracle

1Z0-1158-26

Oracle Cloud Infrastructure Enterprise AI Professional

SHARED BY [u/Jazzy-Pebble-2050](#) on aiexamcerts

EDITION	Free Set 1
QUESTIONS	10
SHARED	October 15, 2026
EXAM	Oracle 1Z0-1158-26, Oracle Cloud Infrastructure Enterprise AI Professional

Shared by [u/Jazzy-Pebble-2050](#) on aiexamcerts

10 practice questions for the Oracle 1Z0-1158-26 exam, Oracle Cloud Infrastructure Enterprise AI Professional, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/oracle/1z0-1158-26. Spotted a mistake? Email support@aiexamcerts.com with the exam code 1Z0-1158-26 and the question number.

1. A support chatbot must give the same answer to the same question across sessions. Which generation setup fits best?

- A High temperature and high top-p for wider variety in wording and tone across replies
- B Maximum output tokens set to the full context window size
- C A high frequency penalty applied to every response
- D Low temperature with a fixed system prompt that defines the answer format

Answer: D

Low temperature reduces sampling randomness, and a fixed system prompt constrains the format, so answers repeat more reliably. High temperature increases variation, a large output limit controls length rather than consistency, and frequency penalties push the model away from repeated tokens.

2. To return the five most similar rows, a query orders by VECTOR_DISTANCE and then limits the result with FETCH FIRST 5 _____.

Answer: ROWS ONLY (also accepted: ROWS)

FETCH FIRST n ROWS ONLY is the standard SQL row limit, and combining it with an ORDER BY on the distance returns the nearest neighbors. Without the ordering, the limit would return an arbitrary five rows.

3. A team needs a model to return answers in a fixed JSON schema. Which prompt design is most reliable?

- A Ask it to be creative and surprise the user with each reply
- B Give an explicit schema, one short example, and an instruction to return only JSON
- C Send the schema once in a separate chat session that is never reused
- D Ask for a long essay and parse only the first sentence

Answer: B

An explicit schema with one worked example and a clear output-only instruction makes structured output far more consistent. Creative instructions invite variation, a separate session loses the constraint, and parsing an essay's first sentence is fragile.

4. Chunking documents into very large pieces always improves retrieval, because each chunk carries more context.

- A True
- B False

Answer: B (False)

Very large chunks dilute the specific passage that matches a query and can crowd out other evidence in the prompt. Chunk size is a tuning decision that should be measured against retrieval quality.

5. Which two controls help a generative AI assistant handle personal data responsibly? (Choose two.)

- A Mask or redact personal data before it is sent to a model
- B Log full prompts that contain personal data into a public bucket
- C Restrict access to the prompt logs with IAM policies
- D Fine-tune on all customer records without any review step
- E Disable the content filters so that responses are returned faster

Answer: A, C

Masking personal data before inference limits what the model and its logs can expose, and IAM restrictions keep stored prompts under access control. Public logs, unreviewed fine-tuning on customer records, and disabled filters each increase risk.

6. Which distance measure is commonly used to compare text embeddings by their direction rather than their magnitude?

- A Hamming distance computed on the raw text bytes
- B Row count of the source table
- C Cosine distance
- D Sequence number of the chunk within its document

Answer: C

Cosine distance compares the angle between two vectors, so it focuses on direction and is insensitive to vector length. Hamming distance on raw bytes, row counts, and sequence numbers do not measure semantic closeness.

7. A RAG answer cites the wrong document. Which part of the pipeline should be checked first?
- A The model's maximum output token setting
 - B The billing alert threshold configured for the entire tenancy account
 - C Chunk metadata and the retrieval ranking for that query
 - D The notebook session shape used during testing

Answer: C

Wrong citations usually come from chunks with bad metadata or from a ranking that puts the wrong passage first, so those are checked first. Token limits, billing thresholds, and notebook shapes do not decide which source is retrieved.

8. Parameter-efficient fine-tuning methods such as _____ add small trainable matrices to a frozen base model.

Answer: LoRA (also accepted: Low-Rank Adaptation)

LoRA, or low-rank adaptation, freezes the base weights and trains small low-rank matrices, which cuts memory and training cost. The adapted behavior can be swapped without copying the full base model.

9. A team has 800 high-quality labeled support transcripts and wants the model to match their tone. Which first step is most appropriate?
- A Train a model from random weights using only the 800 support transcripts available
 - B Prepare clean, representative training pairs and measure a baseline first
 - C Delete the transcripts to avoid any privacy questions
 - D Skip evaluation and deploy the tuned model straight to users

Answer: B

Clean, representative pairs and a baseline measurement show whether tuning actually helps. Training from random weights needs far more data, deleting the transcripts removes the asset needed, and skipping evaluation hides regressions from users.

10. Which element lets an agent call an external function, such as a lookup in an order system?
- A Temperature setting
 - B Embedding dimension count used when the vectors were first generated
 - C Vector index type
 - D Tool or function definition with a parameter schema

Answer: D

A tool or function definition describes the name, purpose, and parameter schema, and the model emits a structured call that the application runs. Temperature, embedding dimensions, and index types control sampling and retrieval rather than external actions.