

PRACTICE QUESTIONS

# SAP Certified - SAP Generative AI Developer

SHARED BY [u/Snappy-Cluster-7362](#) on aiexamcerts

---

EDITION Free Set 1

---

QUESTIONS 10

---

SHARED October 15, 2026

---

EXAM SAP Certified - SAP Generative AI Developer

---

Shared by [u/Snappy-Cluster-7362](#) on aiexamcerts

10 practice questions for the SAP Certified - SAP Generative AI Developer exam, each followed by its answer and a short explanation. More free sets and the full practice exam:

[www.aiexamcerts.com/sap/sap-certified-sap-generative-ai-developer](https://www.aiexamcerts.com/sap/sap-certified-sap-generative-ai-developer). Spotted a mistake? Email [support@aiexamcerts.com](mailto:support@aiexamcerts.com) with the exam name and the question number.

1. Which AI Core concept isolates one tenant's scenarios, configurations and deployments from another tenant's?

- A Service key, which holds the client ID and secret for one user account
- B Resource group
- C Object store secret, which holds credentials for a storage bucket
- D Destination, which stores endpoint details for an application

**Answer: B**

A resource group is the AI Core construct that separates workloads and their objects between tenants or teams. A service key is tempting because it is also tied to an instance, but it carries credentials rather than isolating workloads. Object store secrets and destinations hold connection details and do not partition AI Core resources.

2. In SAP AI Core, a \_\_\_\_\_ references an executable and its parameters, and a deployment runs that object.

**Answer: configuration**

A configuration names the executable and the parameters to run, and deployments are created from it. A resource group is the tempting object to name here, but it isolates workloads rather than defining what runs.

3. Which kind of model converts text into vectors for similarity search?

- A A text-to-speech model
- B A classification model trained on labeled images
- C An embedding model
- D A reinforcement learning policy model

**Answer: C**

An embedding model maps text to numeric vectors so that similar meanings land close together, which is what similarity search needs. Text-to-speech models produce audio and image classifiers label pictures, so they are tempting but unrelated. A reinforcement learning policy chooses actions rather than producing vectors.

4. Temperature is the parameter that controls how random generated text is.
- A True
  - B False

**Answer: A** (True)

Temperature scales the probability distribution over tokens, so higher values produce more varied output and lower values more predictable output. `max_tokens` is a tempting alternative because it also shapes output, but it only caps length. The statement is true.

---

5. Which two steps belong to an ingestion pipeline for grounding? (Choose two.)
- A Rendering the chatbot interface for the user
  - B Creating embeddings for each chunk and storing them
  - C Signing the user's SAML assertion
  - D Replacing the foundation model weights
  - E Splitting documents into chunks

**Answer: B, E**

Ingestion splits documents into chunks and then embeds and stores each chunk for search. The chatbot interface is a tempting step, but it belongs to the application, not the pipeline. Signing assertions and replacing model weights are unrelated to ingestion.

---

6. Why would a team add the masking module to an orchestration pipeline?
- A To compress prompts so they fit the context window
  - B To cache responses for repeated questions
  - C To translate prompts into English automatically
  - D To anonymize or pseudonymize personal data before it reaches the model

**Answer: D**

Masking replaces personal data with placeholders or pseudonyms before the request leaves the controlled pipeline. Compression is a tempting goal because prompts can be long, but masking does not shorten text. Caching and translation are separate concerns with their own mechanisms.

---

7. Which module screens prompts and model responses for harmful or disallowed content?

- A Filtering
- B Translation
- C Templating
- D Masking

**Answer: A**

The filtering module applies content checks to inputs and outputs and can block or flag material above a configured threshold. Masking is a tempting alternative, but it changes personal data rather than judging content. Translation and templating do not evaluate harmfulness.

8. SAP HANA Cloud stores embedding vectors in a column of type \_\_\_\_\_.

**Answer: REAL\_VECTOR** (also accepted: REAL VECTOR)

REAL\_VECTOR is the HANA Cloud column type designed for embeddings, and the similarity functions use it directly. NCLOB is the tempting type for text, but it cannot be compared by vector distance.

9. What is a golden dataset used for in prompt evaluation?

- A Storing live user chat logs for billing purposes
- B Comparing model or prompt versions against expected answers before release
- C Replacing the vector store index during a deployment
- D Training the SAP HANA query optimizer on business data

**Answer: B**

A golden dataset is a curated set of inputs and expected answers used to check each prompt or model version before it ships. Chat logs are a tempting source of test data, but they are not curated expectations and are not for billing. The vector index and query optimizer are unrelated to evaluation sets.

10. Which evaluation practice is most reliable for a generative feature?

- A Read three answers and approve the feature if they sound fluent
- B Measure only the response time of the inference endpoint
- C Score outputs against a fixed set of representative test cases and track regressions over time
- D Count how many tokens each answer uses and compare the totals

**Answer: C**

Fixed, representative test cases give repeatable scores, and tracking them over time shows regressions. Reading a few fluent answers is a tempting shortcut, but fluency does not show correctness. Latency and token counts matter, but they do not measure answer quality.