

PRACTICE QUESTIONS

IABAC

Generative AI Foundation

SHARED BY [u/Solid-Wombat-2318](#) on aiexamcerts

EDITION Free Set 2

QUESTIONS 10

SHARED October 15, 2026

EXAM IABAC Generative AI Foundation

Shared by [u/Solid-Wombat-2318](#) on aiexamcerts

10 practice questions for the IABAC Generative AI Foundation exam, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/iabac/generative-ai-foundation. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. What is a foundation model in the context of generative AI?
- A A model trained only on one narrow task with labeled examples
 - B A model that stores a fixed database of verified answers
 - C A large model trained on broad data and adapted to many downstream tasks
 - D A hardware accelerator designed for training neural networks at data center scale

Answer: C

Foundation models are trained once on broad data and then adapted through prompting, fine-tuning or retrieval for many tasks. A narrow task-specific model is the tempting distractor, since it is also a trained model, but it is built for one job rather than reused across many. A fixed answer database and a hardware accelerator are different things entirely.

2. Generative AI outputs are deterministic, so the same prompt always produces the same response on every system.
- A True
 - B False

Answer: B (False)

Sampling settings such as temperature introduce randomness, and different hardware or model versions can change outputs even at fixed settings. Determinism being the default is the tempting belief. The statement is false.

3. Which sampling parameter limits generation to the smallest set of tokens whose cumulative probability exceeds a threshold?
- A Top-k sampling
 - B Learning rate scheduling
 - C Gradient clipping
 - D Top-p (nucleus) sampling

Answer: D

Top-p, also called nucleus sampling, keeps the smallest set of likely tokens whose probabilities add up to the threshold. Top-k is the tempting distractor, since it also truncates the distribution, but it keeps a fixed number of tokens rather than a probability mass. Learning rate scheduling and gradient clipping affect training, not generation.

4. The number of learned numerical values inside a neural network is called its _____ count.

Answer: parameter

Parameters are the weights and biases learned during training, and their count is a common headline measure of model size. Token count is the tempting answer, since it is also a number quoted for language models, but it measures text length rather than learned values. Context length is the number of tokens the model can read at once.

5. Which two statements about generative video models are correct? (Choose two.)

- A Video generation needs no more compute than a single still image
- B Outputs can be used to create realistic synthetic media that requires disclosure
- C Each frame is always generated independently with no temporal link
- D Video models cannot be connected to text prompts
- E Maintaining consistent objects across frames is a major technical challenge

Answer: B, E

Keeping objects consistent across frames is hard, and realistic synthetic media raises disclosure and misuse concerns. Video needing no extra compute is the tempting distractor, since still images are cheaper, but many frames multiply the cost. Frames are linked temporally in good systems, and text conditioning is common.

6. Which prompt element most directly reduces the risk of a model inventing an answer about a specific internal policy?

- A Raising the temperature so the answer is more creative
- B Asking the model to answer as confidently as possible
- C Removing the policy name from the question so that the model cannot match it to any source
- D Supplying the policy text in the prompt and instructing the model to answer only from it

Answer: D

Providing the source text and constraining answers to it grounds the response in facts the model can check. Raising the temperature is the tempting distractor, since it sounds like it helps with answers, but a higher setting increases invention. Confidence requests and hiding the policy name make fabrication more likely.

7. Which component of a RAG system stores embeddings and returns the nearest matches for a query?
- A A learning rate scheduler
 - B A vector database or vector index
 - C A confusion matrix
 - D A reinforcement learning reward model

Answer: B

A vector database or index stores document embeddings and performs similarity search to return the closest passages. A reward model is the tempting distractor, since it scores outputs in another part of the pipeline, but it does not store documents for retrieval. Learning rate schedulers and confusion matrices are unrelated to lookup.

8. Few-shot examples in a prompt permanently change the model's weights for all future users.
- A True
 - B False

Answer: B (False)

In-context examples only influence a single request, and the model's weights are unchanged once the response is produced. Lasting change being the tempting belief arises from confusing prompting with fine-tuning. The statement is false.

9. What is the main reason a RAG system should cite the source passages it used?
- A So that the model can skip its own training data when it answers any question
 - B So that the embedding dimension is reduced
 - C So that users can verify the answer against the original document
 - D So that the vector index needs no updates

Answer: C

Citations let people check claims against the source, which supports trust and audit. Skipping training data is the tempting distractor, since citing sources can look like a way to bypass the model's knowledge, but the model still generates the answer. Embedding dimensions and index updates are not affected by citations.

10. Which document is most useful for recording a model's intended use, known limitations and evaluation results for stakeholders?

- A A change log of the application's source code
- B A penetration test report for the network
- C A model card
- D A service level agreement with the cloud provider

Answer: C

A model card summarizes purpose, training data, evaluations and limitations in a form that non-engineers can use. A change log is the tempting distractor, since it documents the application's history, but it says nothing about the model's behavior. Network test reports and service agreements serve other purposes.