

PRACTICE QUESTIONS

IABAC

Generative AI Foundation

SHARED BY [u/Patient-Tiger-5493](#) on aiexamcerts

EDITION Free Set 1

QUESTIONS 10

SHARED October 15, 2026

EXAM IABAC Generative AI Foundation

Shared by u/Patient-Tiger-5493 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. Which statement best distinguishes discriminative models from generative models?
 - A Discriminative models always need far more parameters
 - B Discriminative models predict labels from inputs, while generative models learn how the data is produced
 - C Generative models cannot be used for classification at all
 - D Discriminative models only work on text, while generative models only work on images and audio

Answer: B

A discriminative model focuses on predicting a label from the input, while a generative model captures how the inputs are produced and can sample new ones. Saying generative models cannot classify is the tempting distractor, since they are not designed to be discriminators, but they can be used for classification through Bayes rule. Parameter counts and modality limits are not what separates the two families.

2. A model that learns to generate images by pitting a generator against a discriminator is a _____ adversarial network.

Answer: generative

A generative adversarial network trains two networks against each other, one producing images and one judging them. Variational is the tempting answer, since variational autoencoders are another generative family, but they are trained with an encoder and a reconstruction objective rather than a competing discriminator. Recurrent networks process sequences and are not defined by a discriminator.

3. What does a model's context window determine?
 - A How many parameters the model contains
 - B How fast the model is trained on new data during each pretraining run
 - C How many tokens of prompt and conversation fit in one request
 - D How many languages the model supports

Answer: C

The context window is the maximum number of tokens the model can attend to in a single request, including the prompt and the reply. The parameter count is the tempting distractor, since it is also a headline model size figure, but it measures learned weights rather than input length. Training speed and language coverage are separate properties.

4. Increasing a model's temperature setting generally makes its sampled outputs more focused and predictable, while lowering it makes them more varied.
- A True
 - B False

Answer: B (False)

Temperature rescales the probabilities before sampling, so higher values flatten the distribution and produce more varied text, while lower values make outputs more focused. Reversing the direction of the effect is the tempting misreading. The statement is false.

5. Which two measures help limit privacy risks when users send data to a hosted generative AI service? (Choose two.)
- A Reviewing the provider's data retention and training terms
 - B Sending full customer records so the model has more context
 - C Assuming all prompts are deleted immediately without checking
 - D Redacting personal identifiers before the prompt is sent
 - E Turning on every optional data-sharing setting the provider offers

Answer: A, D

Removing identifiers before sending limits what the provider can see, and reviewing retention terms clarifies whether data is kept or used for training. Sending full records is the tempting distractor, since more context seems to improve answers, but it multiplies exposure. Assuming deletion without checking and enabling every data-sharing option are clear mistakes.

6. A customer-support bot is told to answer only questions about the company's products. Which technique does this instruction represent?
- A Gradient clipping during training
 - B Constraining behavior through instructions in the system prompt
 - C Vector quantization of the embeddings
 - D Batch normalization between layers

Answer: B

Scope limits written into the system prompt are a common way to keep an assistant on task. Gradient clipping is the tempting distractor only because it also constrains something, but it limits training updates and has nothing to do with what a deployed bot may answer. Embedding quantization and batch normalization are likewise training or architecture techniques.

7. Which attack tries to override an application's instructions by hiding commands inside user-supplied content?
- A Model extraction through repeated queries
 - B Data poisoning of the pretraining corpus
 - C Prompt injection
 - D Denial of service through large batch sizes

Answer: C

Prompt injection embeds instructions in content such as web pages or documents that the model may follow instead of the developer's rules. Data poisoning is the tempting distractor, since it also manipulates model behavior, but it corrupts training data rather than live prompts. Model extraction and denial of service are different attack families.

8. Numerical vector representations of text that place similar meanings close together are called _____.

Answer: embeddings (also accepted: vectors)

Embeddings map text to vectors so that semantic similarity can be measured, which is how retrieval finds relevant passages. Tokens are the tempting answer, since they are also units of text, but they are discrete pieces rather than dense meaning vectors. Logits are raw output scores for each candidate token.

9. What does retrieval-augmented generation (RAG) add to a language model?
- A Relevant passages fetched from an external knowledge source and included in the prompt
 - B A larger number of parameters learned during pretraining
 - C A new tokenizer trained on the retrieved passages
 - D A guarantee that the model never produces errors

Answer: A

RAG retrieves relevant documents at query time and places them in the prompt so the answer can cite current or private information. More parameters is the tempting distractor, since bigger models seem to know more, but RAG leaves the weights unchanged. A new tokenizer and error guarantees are not provided by retrieval.

10. What does a red-teaming exercise for a generative model aim to do?
- A Rank the model against its competitors on public leaderboards
 - B Probe the system with adversarial inputs before release
 - C Compress the model so it runs on mobile hardware
 - D Estimate the monthly inference cost for each user

Answer: B

Red teaming deliberately tries to elicit unsafe outputs, misuse or policy violations so that gaps can be fixed before release. Leaderboard ranking is the tempting distractor, since it also evaluates the model, but it measures general capability rather than adversarial weakness. Compression and cost estimation do not test for harmful behavior.