

PRACTICE QUESTIONS

IABAC

Certified Artificial Intelligence Expert

SHARED BY [u/golden_pencil_93](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	IABAC Certified Artificial Intelligence Expert

Shared by [u/golden_pencil_93](#) on aiexamcerts

10 practice questions for the IABAC Certified Artificial Intelligence Expert exam, each followed by its answer and a short explanation. More free sets and the full practice exam:

www.aiexamcerts.com/iabac/certified-artificial-intelligence-expert. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. Which layer type is most commonly used to map a sequence of token embeddings to output logits for a vocabulary?
- A A max pooling layer over spatial patches of an image feature map
 - B A layer normalization step applied to the final hidden state
 - C A linear projection followed by a softmax over the vocabulary
 - D An average pooling layer over time steps

Answer: C

A final linear projection produces a score per vocabulary entry, and softmax converts those scores into probabilities. Pooling layers are the tempting distractor, since they appear in many networks, but they reduce spatial size and do not produce a vocabulary distribution. Layer normalization rescales activations but does not produce vocabulary scores.

2. A model trained on one snapshot of a dataset can still be reproduced reliably after the underlying data is overwritten, so data versioning adds little value.
- A True
 - B False

Answer: B (False)

Without versioned data, the exact training inputs are lost, so audits and comparisons become unreliable. Overwriting data being harmless is the tempting belief. The statement is false.

3. What does the dimension of an embedding layer determine?
- A The number of classes in the softmax output layer of the model
 - B The size of each training mini-batch used for this layer
 - C The number of epochs required for convergence
 - D The length of the dense vector representing each discrete item

Answer: D

An embedding dimension sets how many numbers represent each item, such as a token or user id. The output class count is the tempting distractor, since it is also a size on the vocabulary side, but it is a separate layer setting. Batch size and epoch count are training configuration.

4. Training a model on a large, general dataset and then continuing on a smaller task-specific dataset is known as transfer _____.

Answer: learning

Transfer learning reuses general features learned on a broad dataset and adapts them to a narrower task. Curriculum is the tempting answer, since it also orders training data, but it changes the difficulty order rather than reusing weights across tasks. Ensembling combines several models.

5. Which two techniques help reduce overfitting in a deep network? (Choose two.)

- A Increasing the depth of the network without any regularization
- B Applying weight decay as an L2 penalty on the parameters
- C Training on the validation set so that the loss keeps falling
- D Augmenting the data with label-preserving variations of the inputs
- E Removing the held-out evaluation to save compute

Answer: B, D

Weight decay discourages large weights, and label-preserving augmentation expands the variety seen during training. Adding depth without regularization is the tempting distractor, since extra capacity can seem to help, but it usually increases overfitting. Training on validation data and removing evaluation both hide the problem.

6. Which check most directly catches a schema change that would break a model's input pipeline?

- A Increasing the number of training epochs
- B Doubling the batch size on the GPU
- C Validating incoming data against an expected schema and value ranges
- D Renaming the model's output class labels to match the new report format

Answer: C

Schema and range validation fails fast when a column disappears or changes type, before bad inputs reach the model. More epochs are the tempting distractor, since they are a training change, but they do not check incoming data. Batch size and label renaming do not detect input breakage.

7. Which experiment design controls for random variation between runs when comparing two models?

- A Running each configuration once on the largest machine available to the team
- B Repeating each configuration with several seeds and comparing the spread
- C Choosing the seed that gives the best single result
- D Comparing only the training loss curves

Answer: B

Multiple seeds show whether an apparent gain is larger than run-to-run noise. A single run is the tempting distractor, since it is cheaper, but one result cannot separate real improvement from luck. Picking the best seed and comparing only training loss overstate differences.

8. Rotary position embeddings encode relative position by rotating query and key vectors, so attention scores depend on the distance between tokens.

- A True
- B False

Answer: A (True)

Rotary embeddings apply position-dependent rotations to queries and keys, which makes their dot product depend on relative offset. Positions being ignored by attention is the tempting misreading. The statement is true.

9. Which metric is most appropriate for judging a ranking model whose output is a ranked list of items?

- A Mean squared error on every item score
- B Accuracy at a single fixed threshold
- C Normalized discounted cumulative gain
- D Within-cluster sum of squares across all groups

Answer: C

Normalized discounted cumulative gain (NDCG) rewards placing relevant items near the top and discounts lower positions, which matches ranking quality. Mean squared error is the tempting distractor, since it is common for numeric predictions, but it ignores the order of the list. Threshold accuracy and clustering inertia do not evaluate rankings.

10. Which statement about encoder-only transformers is correct?

- A They generate text one token at a time, strictly from left to right
- B They cannot process more than one sentence at once
- C They build bidirectional representations for classification tasks
- D They replace the need for any tokenizer

Answer: C

Encoder-only models attend in both directions, which suits understanding tasks such as classification and embedding. Left-to-right generation is the tempting distractor, since that describes decoder-only models, which are a different family. Sentence limits and tokenizer removal are not properties of encoders.