

PRACTICE QUESTIONS

ARTiBA

Artificial Intelligence Engineer

SHARED BY [u/breezy_turtle_27](#) on aiexamcerts

EDITION Free Set 2

QUESTIONS 10

SHARED October 15, 2026

EXAM ARTiBA Artificial Intelligence Engineer

Shared by [u/breezy_turtle_27](#) on aiexamcerts

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

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

1. A medical screening team wants to catch as many true positive cases as possible, even if that produces extra false alarms that clinicians will review. Which metric should the team emphasize?
 - A Precision
 - B Specificity
 - C Recall
 - D Mean squared error

Answer: C

Recall, calculated as true positives divided by all actual positives, measures how many real cases the screen captures. Precision is the tempting distractor because it measures how many flagged cases are correct, which the team has already agreed to trade away. Specificity concerns true negatives, and mean squared error is a regression loss that does not apply to this screening decision.

2. Cross-validation with k folds produces k trained models whose scores are averaged to estimate generalization performance.
 - A True
 - B False

Answer: A (True)

The statement is true. Each fold serves once as the held-out part while the remaining folds train the model, and averaging the k scores gives a more stable estimate than a single split. Teams often retrain once on all the data afterward for deployment, but that is a separate step.

3. A categorical feature has several thousand distinct values, and one-hot encoding every value would create an enormous sparse matrix. Which approach is most appropriate?
 - A Convert each category into a random integer per row
 - B Group rare categories into an other bucket, or use hashing or target encoding
 - C Drop the column entirely without checking whether it carries any predictive value for the model
 - D One-hot encode every category without any change

Answer: B

Grouping rare levels and using hashing or target encoding keep the feature dimension manageable while preserving most of the signal. Plain one-hot encoding of every value is the tempting default, but it multiplies the number of columns with little benefit. Random per-row integers destroy the meaning of the category, and dropping the column throws away signal without evidence.

4. In pandas, the _____ function reads a CSV file into a DataFrame.

Answer: read_csv (also accepted: `pd.read_csv`, `read_csv()`, `pandas.read_csv`)

The `read_csv` function parses a delimited text file into a DataFrame, and it accepts options for delimiters, column types, and parsing dates. Matplotlib and scikit-learn do not load tabular files, so pandas is the library to reach for during data preparation.

5. Which two statements about batch normalization are correct? (Choose two.)

- A It normalizes activations using mini-batch statistics during training
- B It removes the need to choose a learning rate
- C It learns scale and shift parameters for each normalized feature
- D It is only applicable to image models
- E It uses the statistics of the current batch at inference, so predictions depend on other inputs in that batch

Answer: A, C

Batch normalization normalizes each feature with the mini-batch mean and variance and then applies learned scale and shift parameters. The claim that it removes the learning rate choice is a tempting oversimplification, since the rate still matters. At inference the layer uses running statistics, so predictions do not depend on the other inputs in a batch.

6. Which artifact summarizes a classifier's results as counts of true positives, false positives, true negatives, and false negatives?

- A A confusion matrix
- B A histogram of the target variable
- C The centroids of each cluster found by k-means
- D The magnitude of gradients during each training step

Answer: A

A confusion matrix cross-tabulates actual classes against predicted classes, so precision, recall, and accuracy can all be derived from its four counts. A histogram of the target is the tempting choice, but it shows only one variable and no predictions. Cluster centroids and gradient magnitudes are unrelated to a classification summary.

7. Which ensemble technique reduces variance by training many models on bootstrapped samples and averaging their predictions?
- A Gradient boosting
 - B Bagging
 - C Feature hashing
 - D Early stopping

Answer: B

Bagging, short for bootstrap aggregating, trains independent models on resampled data and averages them, which smooths out the variance of any single model. Gradient boosting is the tempting alternative because it also combines many models, but it trains them sequentially to reduce bias. Feature hashing and early stopping are not ensemble methods.

8. Gradient clipping increases the magnitude of gradients to speed up convergence.
- A True
 - B False

Answer: B (False)

The statement is false. Gradient clipping caps the norm or value of gradients so that updates cannot explode, which is useful in recurrent networks. It limits magnitude rather than increasing it.

9. A company needs an LLM to return JSON that its backend can parse without errors. Which practice most directly supports this?
- A Asking for a friendly tone in the prompt
 - B Increasing the maximum output tokens to 100000
 - C Specifying a schema and validating each response before the backend uses it
 - D Sampling at a temperature of 2.0 with no output schema, so the responses vary freely

Answer: C

A declared schema combined with validation gives the backend a reliable contract and lets malformed responses be rejected or retried. A friendly tone changes style rather than structure, which makes it a tempting but ineffective instruction. A larger output limit and high randomness make malformed output more likely, not less.

10. What is a common reason to fine-tune a model rather than rely on prompting alone?
- A To remove the context window limit completely so that any document fits in one request
 - B To avoid writing any evaluation for the application
 - C To make the model cheaper to host without any change in behavior
 - D To teach a consistent style or domain behavior that prompting does not reliably achieve

Answer: D

Fine-tuning adjusts the weights so a specialized format, vocabulary, or behavior becomes repeatable across requests. Removing the context window limit is a tempting but false promise, because the limit comes from the model's architecture. Fine-tuning does not remove the need for evaluation, and it changes behavior rather than leaving it unchanged.