

PRACTICE QUESTIONS

AI CERTs

AI+ Quality Assurance

SHARED BY [u/Jolly-Gecko-7340](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	AI CERTs AI+ Quality Assurance

Shared by [u/Jolly-Gecko-7340](#) on aiexamcerts

10 practice questions for the AI CERTs AI+ Quality Assurance exam, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/ai-certs/ai-quality-assurance. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. Which statement best describes a golden dataset in AI quality assurance?

- A The most recent production logs, copied into a folder without review or labeling by anyone
- B The vendor's sample data bundled with the model package
- C A curated, labeled set of inputs with agreed expected outputs, reused across releases
- D A synthetic dataset generated to reduce the cost of labeling real records

Answer: C

A golden dataset is curated and labeled against agreed expectations, so the same measurements can be repeated for each release. Raw production logs lack agreed answers and may contain personal data. Vendor samples and synthetic training data have different purposes and no agreed expected outputs.

2. Performance testing is usually the right test type to find defects that appear only under heavy concurrent load.

- A True
- B False

Answer: A (True)

Concurrency, queueing and resource exhaustion problems appear under sustained or peak load, which functional tests rarely generate. Load and stress tests are designed to reproduce those conditions. The distractor is assuming that a passing functional suite covers load behavior.

3. Which level of testing checks that a feature pipeline produces the same values in production as it did in training?

- A Integration testing of the feature pipeline against production-like data
- B Unit testing a single helper function in isolation with mocked dependencies
- C Usability testing of the feature with five end users
- D Load testing the static pages of the public website

Answer: A

Checking pipeline outputs against production-like data exercises the joins, transformations and sources together, which is where training-serving skew appears. A unit test on one helper cannot see the whole pipeline. Usability and load tests answer different questions entirely.

4. Under the EU AI Act, AI systems are sorted into unacceptable, high, limited and minimal risk tiers, and the Act is formally called the EU _____ Act.

Answer: AI (also accepted: Artificial Intelligence)

The regulation is the EU Artificial Intelligence Act, so AI or Artificial Intelligence completes the name. Its risk tiers determine which obligations apply, which is why QA scope depends on classification.

5. Which two facts should a defect report state about a model output error? (Choose two.)

- A The exact input or prompt that produced the output
- B The name of the developer who wrote the feature
- C The model, prompt or configuration version used when it occurred
- D The screen resolution of the monitor used during testing
- E The number of test sessions the tester ran that week

Answer: A, C

Reproducing a model error needs the exact input and the version of the model, prompt and configuration in use, since any of them can change the output. Staff identifiers, hardware color and break counts do not help a developer reproduce anything. Focus the report on the variables that determine the output.

6. Which statement correctly distinguishes severity from priority?

- A Severity and priority are always the same value in every project
- B Priority describes the number of users affected, while severity describes the cost
- C Severity is set by the customer who files the ticket, while priority is set by the developer who owns the code
- D Severity describes the technical impact of a defect, while priority describes how soon it should be fixed

Answer: D

Severity measures how badly the defect affects the system, and priority sets the scheduling order based on business need. Treating them as identical loses the ability to fix a cosmetic issue on a critical date. The option that assigns the two fields to users and costs swaps their meanings.

7. Which record shows that a bias test was actually run before a release?

- A A verbal assurance at the stand-up from the data scientist that the bias results were fine
- B A dated evaluation report listing groups tested, metrics used and results
- C A link to the vendor brochure about its responsible AI program
- D The number of meetings the test team held during the sprint

Answer: B

An evaluation report with dates, groups, metrics and results is evidence an auditor can check. A verbal assurance or marketing link cannot be verified after the fact. Meeting minutes count events without showing any measurement.

8. A defect escape rate should be read alongside the number of defects found before release.
- A True
 - B False

Answer: A (True)

Escape rate only has meaning as a ratio of defects found after release to those found overall, so the pre-release count is part of the picture. Reading it on its own can make a team look better or worse than it is. The distractor is treating the escape number as a standalone score.

9. Which pytest feature runs one test function with several sets of input values?
- A A session-scoped fixture, so setup runs once for the whole test session
 - B The conftest.py file that holds a skip hook
 - C The parametrize marker, `pytest.mark.parametrize`
 - D The xfail marker with a reason string

Answer: C

The parametrize marker generates one test case per input set, keeping the test body in one place. Session-scoped fixtures change how setup is shared rather than how inputs vary. The xfail marker expects a failure, which is a different concern from data-driven testing.

10. A feature has five parameters with three values each. Which technique covers every pair of values with far fewer than the 243 full combinations?
- A Pairwise, or all-pairs, testing
 - B Exploratory testing with session charters
 - C Mutation testing with a minimum score threshold
 - D Equivalence partitioning applied to each single field

Answer: A

Pairwise design covers every two-way interaction with a small set of combinations, which is effective because most defects involve only a few parameters. Exploratory and mutation testing are valuable but do not produce a systematic combination set. Single-field partitioning ignores interactions between fields entirely.