PRACTICE QUESTIONS

ISTQB

CT-AI

Certified Tester AI Testing

PUBLISHED BY aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
ISSUED	November 8, 2026
EXAM	ISTQB CT-AI, Certified Tester AI Testing

Shared by **aiexamcerts** on aiexamcerts

10 practice questions for the ISTQB CT-AI exam, Certified Tester AI Testing, each followed by its answer and a short explanation. More free sets and the full practice exam: **aiexamcerts.com/istqb/ct-ai**. Spotted a mistake? Email support@aiexamcerts.com with the exam code CT-AI and the question number.

---

1. Acceptance for an autonomous car states that it must stop within a maximum distance after an obstacle is detected. Which approach to the test oracle problem does this illustrate?

A Defining output boundaries, such as acceptable ranges and tolerances
B Proxy oracles using a secondary model
C Expert consultation to label each result
D Back-to-back testing against an older version

Answer: A

Defining output boundaries means agreeing limits, ranges or tolerances for acceptable results instead of an exact expected value. A proxy oracle would instead use another system or model to judge the output.

---

2. Back-to-back testing can compare an operational data pipeline against the training pipeline to verify that they provide consistent functionality.

A True
B False

Answer: A (True)

The syllabus lists back-to-back testing in data pipeline system testing for exactly this purpose, and again in production to compare against previous versions.

---

3. For which fine-tuning scenario would success be most likely, based on the similarity between the original and the new task?

A Adapting the cat breed image classifier to recognize spoken accents
B Adapting a speech recognizer to predict stock prices
C Adapting a dog breed classifier to detect fraudulent transactions
D Adapting a cat breed image classifier to recognize dog breeds

Answer: D

Fine-tuning success depends on how similar the new task is to the original one, and cat breeds to dog breeds is a small change. Spoken accents differ much more from images of animals, so fine-tuning would be less effective.

---

4. A team compares a conventional spam filter with an ML-based one. Which statement describes the fundamental difference in how the two are built?

A The conventional filter infers its rules from labeled emails, while the ML-based filter is programmed with step-by-step if-then-else logic

B The ML-based filter derives its behavior from patterns in training data, while the conventional filter follows rules that developers wrote explicitly

C Both filters learn from data, but only the conventional filter can produce probabilistic outputs for the same input

D The ML-based filter always behaves deterministically, so its outputs are easier to trace than those of the conventional filter

Answer: B

Conventional systems are programmed imperatively with explicit instructions, whereas most ML-based systems analyze data to learn patterns that they then apply to new inputs. The option claiming the ML-based filter is always deterministic is wrong: probabilistic reasoning and limited explainability are typical of AI-based systems.

---

5. When testing a GenAI model, a parameter that influences the randomness of the generated output, alongside maximum tokens, is the _____.

Answer: temperature

The syllabus names temperature and maximum tokens as parameters that affect GenAI output, and they add to the input explosion problem. Together with the context window they make the same prompt produce different results.

---

6. In a dataset, each row has a `start_date` and an `end_date`, and `end_date` must always be later than `start_date`. Which constraint test is used?

A A single-value Type constraint on each date

B A comparison constraint of the Greater Than kind

C A multi-value Sum constraint across all rows

D A multi-value Outlier constraint on the dates

Answer: B

A comparison constraint compares values of different attributes, and Greater Than tests that one attribute value exceeds another. A Type constraint would only check that each value is a date.

---

7. A defect prediction model produced these confusion matrix results on a test set: TP = 45, FP = 5, FN = 15, TN = 135. What is its recall?

A 75%

B 90%

C 96.4%

D 25%

Answer: A

Recall = TP / (TP + FN) = 45 / 60 = 75%. The value 90% is the precision (45 / 50) and also the accuracy (180 / 200), so it is a tempting but wrong choice for recall.

---

8. In reinforcement learning, the agent learns from a labeled dataset that contains the correct output for every input.

A True

B False

Answer: B (False)

Reinforcement learning uses rewards and penalties from interaction with the environment, not a labeled dataset. Learning from labeled input-output pairs is supervised learning.

---

9. Which description best fits red teaming of an AI-based system?

A A comparison of two model variants on live traffic to find which performs better

B Continuous real-time filtering of inputs to an operational system

C A systematic, often black-box fault attack that probes the system to find harmful capabilities, particularly in its outputs

D A review of the model documentation against a checklist

Answer: C

Red teaming deliberately attempts to make the system produce harmful or undesirable results so that these capabilities can be mitigated. The real-time defensive monitoring and filtering of inputs describes blue teaming, which complements red teaming.

---

10. Training and validation errors both remain high and close together on the learning curves, with little improvement as training progresses. What does this indicate?

A Overfitting

B Data poisoning

C Reward hacking

D Underfitting

Answer: D

Consistently low metrics on training and validation sets, shown as high and close errors on the learning curves, indicate an underfit model. In overfitting, training performance is good while performance on unseen data is poor.