ISTQB Certified Tester
AI Testing v2.0
Sample Exam Answers

American Software Testing Qualifications Board

40 questions | 44 points possible | 30 points needed to pass

1. (1 point) Which of the following is the best description of the nature of an AI-based system?

a. Black-box
b. Gray-box
c. White-box
d. Clear-box

A is correct. AI-1.1.1 (K2) Differentiate between AI-based systems and conventional systems

2. (1 point) What type of AI is used in the vast majority of systems currently in use?

a. Narrow AI
b. Frontier AI
c. General AI
d. Super AI

A is correct. AI-1.1.2 (K2) Distinguish between narrow AI, general AI, and super AI

3. (1 point) How does reinforcement learning work in ML (machine learning)?

a. It utilizes labeled data and algorithms
b. It looks for patterns in unlabeled data
c. It uses trial-and-error interactions to learn optimal behavior
d. It uses decision trees for data classification

C is correct. A and D are referring to supervised learning. B is referring to unsupervised learning. AI-1.1.3 (K2) Explain the different types of AI technologies

4. (1 point) What is one of the biggest problems with GenAI?

a. It cannot create realistic text yet
b. Pictures and videos of humans it generates make people look robotic
c. There are significant ethical concerns and potential societal issues
d. They can only be used by highly technical people, reducing availability

C is correct. A and B are issues, but not the biggest problem.
D is not true because everyone is using AI now. AI-1.1.4 (K2) Explain generative AI

5. (1 point) What is the purpose of a neuromorphic processor?

a. It is used for AI and has an architecture that loosely resembles the way brain neurons are structured
b. It is used to kill AI processes that have run out of control by shutting down parts of their neural networks
c. It is used for running traditional operating systems more efficiently
d. It is used to store large sets of data that are needed for GenAI text development

A is correct. AI-1.1.5 (K2) Compare the choices available for hardware to implement machine learning systems

6. (1 point) Which of the following is a problem with public cloud hosting of an AI system?

a. Speed
b. Processing power and suitability
c. Data privacy and security
d. High infrastructure costs

C is correct. Data privacy and security are the biggest concerns with public cloud platforms. A and B are benefits from the scalability of the systems. D is a problem with on-premises systems where power and cooling are big concerns. AI-1.1.6 (K2) Compare the options for the development and hosting of AI models

7. (1 point) What are the two perspectives of quality for AI-based systems as per ISO 25059?

a. Project quality and product quality
b. Quality in use and resistance to abuse
c. System quality and user quality
d. Product quality and quality in use

D is correct. AI-2.1.1 (K2) Classify behaviors of AI-based systems according to the quality characteristics defined in ISO/IEC 25059

8. (1 point) What safety-related challenge is being encountered when sufficient safety standards are not available to use when a system is being developed?

a. Non-determinism
b. Self-learning
c. Explainability and transparency
d. Evolving regulations

D is correct. In this case, the safety-standards are either not available or are changing, so the system can't be created in accordance with the standards. AI-2.1.2 (K2) Explain the special considerations that arise when AI is used in safety-related systems

9. (1 point) You are working with an AI-based system that controls drug trial scheduling for new drugs. You have defined the following acceptance criteria:

The system must provide sufficient data to meet the FDA requirements for how people are assigned to cohorts to take the drug under test or a placebo

This is an example of an acceptance criterion for which quality characteristic?

a. Functional correctness
b. User controllability
c. Functional adaptability
d. Transparency

D is correct. AI-2.2.1 (K2) Give examples of acceptance criteria for AI-based systems

10. (1 point) In machine learning, regression is used with which type of learning?

a. Classification
b. Supervised
c. Unsupervised
d. Reinforcement

B is correct. Regression is used in supervised learning in predicting continuous numerical values. AI-3.1.1 (K2) Distinguish between the different forms of ML

11. (1 point) What is the first step in the ML workflow?

a. Select a Framework
b. Select and Build the algorithm
c. Prepare and Test Data
d. Understand the Objectives

D is correct. AI-3.1.2 (K2) Summarize the workflow used to create an ML system

12. (1 point) Which of the following lists the key activities in data preparation?

a. Data acquisition, data pre-processing, feature engineering
b. Data creation, data updates, data deposits
c. Data extraction, data transformation, data loading
d. Data addition, data modification, data deletion

A is correct. AI-3.2.1 (K2) Explain the activities related to data preparation

13. (1 point) Why is the test dataset referred to as the "holdout dataset"?

a. Because the data is not usually representative of the production data, but is designed for testing corner cases
b. Because the data will be used for the lower-priority negative test cases, and these are "held out" of execution in the event of short timelines
c. Because the dataset is not used for training or evaluation, it is "held out" for just testing the tuned model
d. Because the dataset includes data that must be protected due to privacy laws, and anonymization is required before it can be used

C is correct. AI-3.2.3 (K2) Contrast the use of training, validation, and test datasets in the development of an ML model

14. (1 point) How are neurons in one layer connected to the other neurons in a fully connected, standard neural network?

a. Each neuron is connected to all other neurons in the neighboring networks (friendly connection)
b. Each neuron in a layer is connected to all other neurons in the next layer (one-way connection)
c. Each neuron in a layer is connected to all the input and output neurons
d. Neurons do not connect to each other but rather connect to neuroids that handle all the communication between the neurons

B is correct. AI-3.4.1 (K2) Explain the structure and working of a deep neural network

15. (1 point) What type of neural network structural coverage measurement measures the proportion of neurons where their output either exceeds the maximum output achieved during training, or is less than the minimum output achieved during training?

a. Neuron coverage
b. K-Multisection Neuron Coverage (kMNC)
c. Neuron Boundary Coverage (NBC)
d. Neuron Activation and Bias Coverage (ABC)

C is correct. This testing is to see if the responses are over the boundaries that were established during training. AI-3.4.3 (K2) Describe the different coverage measures for neural networks

16. (2 points) You have been given the following confusion matrix:

		Actual	
		Positive	**Negative**
Predicted	**Positive**	True Positive 50	False Positive 25
	Negative	False Negative 25	True Negative 50

Given this information, what is the Precision of the model's predictions?

a. 50%
b. 66.7%
c. 87.5%
d. 100%

B is correct. The formula is TP / (TP + FP) * 100%, which is (50) / (50 + 25), which is 66.66667 percent. AI-3.3.1 (K3) Calculate common ML functional performance metrics from a given set of confusion matrix data

17. (1 point) Why is the maintenance testing of an adaptive AI-based system more difficult to automate?

a. Because user usage will vary over the life of the system
b. Because input data may become biased
c. Because it is deterministic and will not change its behavior after deployment
d. Because it will continue to learn and change after it has been deployed

D is correct. An adaptive system will continue to change as it learns and may not produce the same outputs for the same inputs as time goes on. AI-4.1.1 (K2) Compare the testability of locked and adaptive AI-based systems

18. (1 point) Which of the following is an example of why a statistical approach is needed during testing?

a. The model incorrectly classified a rabbit as a rhinoceros one time
b. The model is running very slowly when processing text
c. The model is producing incorrect outputs 100% of the time
d. The model is acting consistently in a deterministic manner

A is correct. Since this happened only one time, statistics are needed to determine if this is a significant issue in the responses. AI-4.1.2 (K2) Explain why a statistical approach is often needed when testing AI-based systems

19. (1 point) You have just deployed a product that helps a user select clothing based on their preferences and previous choices. This system learns based on selections and rejections from the user as well as style trends. Why is it difficult to define a test oracle for this system?

a. Because the specifications are incomplete
b. Because there is subjectivity in determining the correct behavior
c. Because the system is deterministic
d. Because the input boundaries have not been defined

B is correct. The problem with the test oracle is that it also has to consider a user's individual taste, and what they like or dislike is subjective. AI-4.1.3 (K2) Explain the challenges and solutions relating to test oracles for AI-based systems

20. (1 point) How are GenAI outputs usually evaluated?

a. Correctness, coherence, creativity
b. Copyright, control, comprehension
c. Prompts, pictures, parameters
d. People, processes, products

A is correct. AI-4.2.1 (K2) Explain how generative AI can be tested

21. (2 points) You have been conducting a red teaming exercise on a GenAI chatbot that advises university candidates which majors are most appropriate for their skills and interests. The team has discovered that there are certain sets of inputs that result in students being told they should not pursue an education because they are too stupid.

What should you do now?

a. Terminate all work on the product and start over since something has gone very wrong
b. Form a new team and conduct the testing again, but this time from the faculty perspective
c. Apply the inputs the team used and train or modify the system to not produce harmful outputs
d. Form a blue team to conduct the same tests so a comparison can be made

C is correct. The information used by the team to create the harmful output should be used to modify or train the system to provide more acceptable responses. A is not correct because this is probably not a feasible approach, and the system is not useless, just needs modification. B is not correct because the student perspective is the right approach, and the red team has been successful in finding issues. D is not correct because blue teams conduct testing in production to protect against attacks. AI-4.2.2 (K3) Implement red teaming for GenAI systems

22. (1 point) What data flow is tested during input data testing?

a. The data generated by the ML model
b. The data that has been extracted prior to applying feature engineering
c. The data that is used to train the model and the data that is used in production
d. The data that has been masked or anonymized to ensure integrity is unaffected

C is correct. AI-4.3.1 (K2) Summarize the test levels used to develop machine learning systems

23. (1 point) What are the three main areas used to categorize risk using the ML workflow?

a. Design, Analysis, Development
b. Selection, Model testing, Data control
c. Development, Input data, Model
d. Data, Model, Algorithm

C is correct per the syllabus. AI-4.3.2 (K2) Explain how risk-based testing is applied to machine learning systems

24. (1 point) Which of the following activities would be best to use to mitigate a risk from wrong data types in the training data?

a. A/B testing
b. Testing for bias
c. Data constraint testing
d. Data representativeness testing

C is correct. In this case, it's likely that the constraints were not sufficient when the data was created/generated. The constraints should be tested. AI-5.1.1 (K2) Give examples of test approaches used for the risk mitigation of input data for a machine learning system

25. (1 point) What are the primary sources of bias in an AI-based system?

a. Training data and defects in the algorithm or model
b. Disparate impact analysis and design flaws
c. Incorrect labelling and invalid boundary declarations
d. Poor documentation and security vulnerabilities

A is correct. These are the most likely sources of bias. AI-5.1.2 (K2) Explain how to test for bias

26. (1 point) How is back-to-back testing used in data pipeline testing?

a. To ensure the correct pipeline code versions
b. To verify compliance with the test strategy
c. To test the pipeline's compliance with specified requirements
d. To verify that the performance is consistent or better than previous versions

D is correct. Back-to-back testing tests different versions and compares the results. AI-5.1.3 (K2) Summarize the various forms of data pipeline testing

27. (1 point) What are the typical three steps for testing for data representativeness?

a. Define the target population, analyze data characteristics, and apply statistical assessment techniques
b. Understand the intended use cases, analyze the characteristics of the end users and operational environments, and identify the expected operation data distributions
c. Apply EDA to the training and reference datasets, visualize distributions, and identify potential anomalies
d. Use formal statistical tests to compare distributions, check for data imbalances, and verify adequate coverage for the typical and edge scenarios

A is correct. B is not correct; these are the steps for defining the target population. C is not correct; these are the steps for analyzing data characteristics. D is not correct; these are the steps for applying statistical assessment techniques. AI-5.1.4 (K2) Explain how to test for data representativeness

28. (2 points) You are planning to perform some dataset constraint testing. You have a situation where the value of one attribute, income, must be greater than another attribute, taxes paid. You want to make a comparison of all the data to ensure that these attributes have the proper relationship.

What type of testing should you apply?

a. Outlier
b. Correlate
c. Greater Than
d. Sum

C is correct. This is done to check that one value is greater than another. AI-5.1.5 (K3) Apply dataset constraint testing (2 pts)

29. (1 point) You want to use an expert review to verify the accuracy and consistency of the labels in your supervised learning model. Which experts should you use?

a. Statisticians
b. Developers
c. Testers
d. Trained annotators

D is correct. Expert reviews are conducted by domain experts or trained annotators. AI-5.1.6 (K2) Explain label correctness testing

30. (1 point) Your recently deployed AI-based resume screening system is rejecting all resumes from men, even though this is not a job requirement. What type of testing should be conducted to catch this type of problem?

a. Bias testing
b. Ethical system testing
c. Overfitting
d. Reward hacking

A is correct. This appears to be a bias problem, so testing for bias should be able to detect this problem. AI-6.1.1 (K2) Give examples of test approaches used for risk mitigation of ML models

31. (1 point) Why is comprehensive documentation of the ML model critical as compared to documentation for a traditional system?

a. Because white-box testing of the model is very expensive
b. Because the system contains machine-generated code and is generally considered a black-box
c. Because the data analysts need to assess the validity of the data to be used by the system
d. Because the operations people need to have a detailed description of the model's calculations to plan deployment

B is correct. The documentation is needed because the system itself is hard to understand and cannot be analyzed sufficiently to determine how it is working. AI-6.1.2 (K2) Explain the purpose and focus of reviewing ML model documentation

32. (1 point) Why are single pass/fail checks not sufficient for probabilistic ML systems?

a. Because proper behavior can only be determined across a large, representative dataset
b. Because any one failure should result in the system being retrained
c. Because at least two pass results are required before a test case can be confirmed
d. Because this will not result in sufficient defect reports to justify the testing effort

A is correct. Because the results won't always be right, the overall performance of a system must be determined by using a large dataset of representative data. AI-6.1.3 (K2) Explain how ML functional performance testing is carried out for probabilistic machine learning systems

33. (1 point) What is an adversarial example?

a. It is an input that causes the model to make incorrect predictions
b. It is a valid input that is processed incorrectly by the model
c. It is a form of security vulnerability that allows data to be accessed and sold to adversaries
d. It is a form of blue teaming that measures how well a system can detect and defend against an attack

A is correct. AI-6.1.4 (K2) Summarize adversarial testing of machine learning systems

34. (2 points) You are testing an AI-based system that predicts the width of tree rings based on the annual rainfall. Research has shown that higher rainfall produces wider rings, and, conversely, lower rainfall produces tighter rings.

Given this information, which of the following is the correct application of metamorphic testing?

a. Vary the tree ring width and see if the rainfall amount is adjusted accordingly
b. Vary the rainfall amount and see if the predicted tree ring width has changed
c. Define a set tree ring width and vary the rainfall amount until that width is predicted
d. Define the rainfall amount as a constant and vary the other parameters that affect the tree ring width

B is correct. The relationship is between the rainfall and the tree ring width, in that when the rainfall increases, the width gets bigger and vice versa. AI-6.1.5 (K3) Use metamorphic testing to derive test cases for a given scenario

35. (1 point) What type of drift testing depends on the current ground truth being known?

a. Concept
b. Data
c. Dynamic
d. Static

C is correct. Dynamic drift testing requires the current ground truth, which can then be compared to the model's output to determine if there is a difference between the two. This difference is then measured to see if it matters. AI-6.1.7 (K2) Explain how drift testing is used on operational machine learning systems

36. (1 point) You have been testing a model, and you are seeing that the functional performance metrics are much worse than expected. This is an indication of what type of problem?

a. Overfitting
b. Underfitting
c. Right-fitting
d. Sub-fitting

B is correct. This is an indication of underfitting where the model was trained on data that was too simple, so that it doesn't perform well with the validation and test data. AI-6.1.8 (K2) Explain how overfitting and underfitting are detected by testing

37. (1 point) How can A/B testing be applied to a production AI-based system when a change is made to the system?

a. By comparing the new system's functional performance metrics to the metrics from the version before the change
b. By comparing the new system's outputs to the pseudo-oracle
c. By generating automated tests that can perform a complete regression test of all functionality
d. By generating use cases based on normal usage and executing those against the new version

A is correct, this is A/B testing. B is talking about back-to-back testing. C and D are just not correct as nothing is generated automatically in A/B testing. AI-6.1.9 (K2) Explain how A/B testing is used in the context of machine learning systems

38. (1 point) How should back-to-back testing be used for an AI-based system?

a. By providing the same input to the existing system and the new version of the system, and comparing the functional performance metrics
b. By providing the same input to the new system and the pseudo-oracle and comparing the outputs
c. By determining a relationship between the inputs and outputs and varying the inputs to see if the outputs vary accordingly
d. By determining the edges and boundaries and developing test input data that will cause the system to provide outputs on or over those boundaries

B is correct. This is an example of back-to-back testing. A is A/B testing. C is metamorphic testing. D is boundary value analysis. AI-6.1.10 (K2) Explain how back-to-back testing is used in the context of machine learning systems

39. (1 point) Which of the following is the best type of testing to detect if the evaluation approach is sub-optimal?

a. A/B testing
b. Smoke testing
c. Functional performance testing
d. Performance efficiency testing

C is correct. It's best to check the functional performance (accuracy, etc.) to see if the evaluation of the outputs was good. AI-7.1.1 (K2) Give examples of test approaches used for risk mitigation of ML development

40. (1 point) How does rollback testing work for a new deployment?

a. It verifies the system's capability to revert to a previously stable version
b. It runs the new model alongside the current production model in real-time and compares the results
c. It ensures that the new model still meets the required performance efficiency in the new environment
d. It validates that a small subset of the production traffic can be processed by the new model without degradation

A is correct. Rollback testing ensures that the system can be rolled back to a previous, stable version in the case of a failure of the new version. AI-7.1.2 (K2) Explain the various forms of ML system deployment testing