Sample Exam – Questions

Sample Exam set A
Version 2.0

ISTQB® AI Testing Syllabus

Compatible with Syllabus version 2.0

International Software Testing Qualifications Board

Version 2.0 Released 2026/04/17

Copyright Notice

Document Responsibility

The ISTQB® Examination Working Group is responsible for this document.

This document is maintained by a core team from ISTQB® consisting of the Syllabus Working Group and Exam Working Group.

Acknowledgements

This document was produced by a core team from ISTQB®: Klaudia Dussa-Zieger, Stuart Reid, Vipul Koch, Kyle Siemens, Qin Liu, Werner Henschelchen, Jarosław Hryszko

The core team thanks the Exam Working Group review team, the Syllabus Working Group and Member Boards for their suggestions and input.

Revision History

Sample Exam – Questions Layout Template used:	Version 2.12	Date: November 27, 2024

Version	Date	Remarks
1.0	2021/10/01	Release for GA
2.0 Beta	2026/01/05	Beta Review
2.0	2026/04/17	Release for GA

Table of Contents

Introduction

Purpose of this document

The example questions and answers and associated justifications in this sample exam have been created by a team of subject matter experts and experienced question writers with the aim of:

- Assisting ISTQB® Member Boards and Exam Boards in their question writing activities
- Providing training providers and exam candidates with examples of exam questions

These questions cannot be used as-is in any official examination.

Note, that real exams may include a wide variety of questions, and this sample exam ***is not*** intended to include examples of all possible question types, styles or lengths, also this sample exam may both be more difficult or less difficult than any official exam.

Instructions

In this document you may find:

- Questions[1], including for each question:
  - Any scenario needed by the question stem
  - Point value
  - Response (answer) option set
- Additional questions, including for each question [does not apply to all sample exams]:
  - Any scenario needed by the question stem
  - Point value
  - Response (answer) option set

- *Answers, including justification are contained in a separate document*

[1] In this sample exam the questions are sorted by the LO they target; this cannot be expected of a live exam.

Questions

Question #1 (1 Point)

Which of the following statements BEST highlights the difference between conventional and AI-based systems?

a) AI-based systems tend to learn from patterns in data and can adapt to new scenarios, while conventional systems follow predefined rules and produce consistent outputs for the same inputs

b) AI-based systems tend to be faster than conventional systems because they do not rely on explicit programming, while conventional systems are slower due to rigid rule-based processing

c) AI-based systems tend to be deterministic and explainable, while conventional systems are probabilistic and often difficult to interpret

d) AI-based systems tend to be more suitable for critical tasks where explainability is crucial, while conventional systems are better at handling complex tasks in unregulated areas

Select ONE answer.

Question #2 (1 Point)

Which TWO of the following statements **BEST** distinguish between narrow AI, general AI, and super AI?

a) Narrow AI is self-learning, general AI focuses on solving specialized problems, and super AI is limited to tasks defined during its development

b) Narrow AI operates independently of human input, general AI is used only in advanced robotics, and super AI improves human decision-making in specialized fields

c) Narrow AI performs specific tasks efficiently, general AI can handle a wide range of intellectual tasks like a human, and super AI surpasses human intelligence

d) Narrow AI is task-specific, general AI is a concept with limited real-world applications, and super AI describes cutting-edge generative AI models

e) Narrow AI is the current state of AI, achievement of general AI is uncertain, but once general AI is achieved, super AI is likely to happen

Select TWO answers.

Question #3 (1 Point)

Which of the following statements **BEST** describes the relationship between AI, ML, and DL?

a) ML is a subset of AI, and DL is a further subset of ML, representing a hierarchy of specialized technologies

b) DL is the foundational technology, and both ML and AI are specialized applications built upon its principles.

c) DL and ML are essentially interchangeable terms, while AI represents a separate approach to problem-solving

d) AI encompasses both ML and DL as distinct but important methodologies, working in parallel to solve complex problems

Select ONE answer.

Question #4 (1 Point)

Which of the following statements **BEST** explains generative AI?

a) It is designed to analyze and understand existing data, such as text or images, allowing machines to categorize information and extract key insights

b) It creates new content, such as text, images, or audio, by learning patterns from training data and generating outputs similar in nature

c) It focuses mainly on improving existing content, such as text, images, or audio, by optimizing it for classification or prediction tasks

d) It creates models that are limited to generating text and images and cannot be used for creative tasks like drug discovery or data simulation

Select ONE answer.

Question #5 (1 Point)

Which of the following statements **BEST** compares the hardware options available for implementing machine learning systems?

a) GPUs are well-suited for training and running ML models due to their ability to handle massively parallel processing, while CPUs are better for general-purpose computing

b) CPUs are more efficient than GPUs for ML tasks because they have faster clock speeds and can handle complex operations

c) AI-specific hardware, such as ASICs, is primarily used for training ML models, while GPUs are better suited for edge computing

d) Neuromorphic processors are a form of AI hardware specifically optimized for running on the von Neumann architecture

Select ONE answer

Question #6 (1 Point)

Given the following statements about AI model development and hosting:

i. It achieves lower development costs by using public cloud resources, eliminating the need for local hardware investment

ii. It uses local development of data preparation components for sensitive data for increased security before moving to the cloud for training of the full system

iii. It results in lower costs because laptops are used for local development and there are low upfront hardware costs by hosting AI models on public clouds

iv. It simplifies development and hosting by standardizing processes on local servers, removing the need for complex cloud-based configurations

v. It guarantees the highest security by hosting AI models on private clouds, thereby avoiding the risks associated with local hardware vulnerabilities

Which of the following **BEST** reflects advantages of using a hybrid approach for this development and hosting?

a) i, ii and v

b) ii and iii

c) iii and iv

d) i, iv and v

Select ONE answer.

Question #7 (1 Point)

Given the following ISO/IEC 25059 quality characteristics:

1. AI Robustness
2. User controllability
3. Functional adaptability
4. Intervenability

And the following examples of quality characteristic measures:

A. The success rate of a remote operator in forcing a drone into the safe-landing protocol when its AI navigation system exhibits hazardous behavior

B. The average time required to successfully override a fraud management system's automated decision to block a customer's transaction

C. The F1-score of an object detection model in an autonomous car in heavy rain

D. The time required for an e-commerce recommendation engine to update its suggestions to reflect a new, rapidly emerging fashion trend

Which of the following **BEST** matches the quality characteristics with the example measures?

a) 1D – 2A – 3C – 4B

b) 1C – 2D – 3B – 4A

c) 1C – 2B – 3D – 4A

d) 1A – 2D – 3C – 4B

220 Select ONE answer.
221

222

Question #8 (1 Point)

Which of the following statements **BEST** describes a key challenge when AI is used in safety-related systems?

a) When the requirements are too detailed, it leaves little room for the ML system to learn from the implicit goals contained within the training data

b) The potential for non-determinism and self-learning makes some AI-based systems unpredictable as they diverge from their original tested state

c) Because self-learning safety-related systems stop adapting after deployment, safety is often compromised by the need for manual updates of the operational AI-based system

d) The mature safety-related standards tend to be out-of-date and require the use of outdated AI technologies, which hinders innovative AI solutions in AI-based systems

Select ONE answer.

Question #9 (1 Point)

Which of the following examples is **LEAST** likely to be a valid acceptance criterion for AI-specific quality characteristics defined in the ISO 25059 standard for an AI-based system?

a) The security guard in the museum control room can trigger an immediate 'all-stop' command that causes the patrol robot to cease all motion within 0.5 seconds to prevent collision with a sculpture

b) A greenhouse control system reacts within 20 minutes when the measured humidity is greater than 10% from the optimum humidity

c) The spam alert control system is easy for the user to set up and requires minimal technical expertise to maintain

d) When the analysis tool flags a retinal scan for severe diabetic retinopathy, it shall display a visual heatmap overlay on the image, highlighting the key features

Select ONE answer.

Question #10 (1 Point)

Given the following forms of ML:

1. Clustering
2. Reinforcement learning
3. Classification
4. Regression

And the following examples:

A. The mobile game app updates its feedback, response timing, and the number of user options it provides based on how much the players spend
B. The language translation app searches the internet to find text provided in multiple languages to improve its translation function
C. A manufacturing company predicts when equipment is likely to fail based on sensor data and historical maintenance records
D. A social network platform groups its users into communities based on their interactions with each other and their stated interests

Which of the following **BEST** matches the examples with the forms of ML?

a) 1B – 2C – 3D – 4A
b) 1A – 2D – 3B – 4C
c) 1C – 2D – 3A – 4B
d) 1D – 2C – 3B – 4A

Select ONE answer.

Question #11 (1 Point)

Given the following activities from the ML workflow:

1. Deploy the Model
2. Prepare & Test Data
3. Test the Model
4. Evaluate the Model

And, given the following descriptions:

A. Model performance is tested using validation data
B. The origin of the test data used to test the model is identified
C. Test data are used to verify the agreed performance criteria are met
D. The model is tested on the target platform

Which of the following **BEST** matches the descriptions with the activities in the ML workflow?

a) 1A – 2C – 3B – 4D
b) 1C – 2D – 3B – 4A
c) 1D – 2B – 3A – 4C
d) 1D – 2B – 3C – 4A

Select ONE answer.

Question #12 (1 Point)

Which of the following statements about the use of pre-trained models is **CORRECT**?

a) When applied to a neural network, the RAG approach works by adding additional layers that hold documentation specifically related to the prompt
b) Fine-tuning a biased LLM with high-quality, task-specific data prevents unfair outputs based on sensitive attributes
c) To successfully adapt a pre-trained neural network, fine-tuning requires that additional training with new data is applied to all layers of the network
d) The RAG approach requires the identification and acquisition of data relevant to the task up front, but requires no changes to the pre-trained model

Select ONE answer.

Question #13 (1 Point)

Which of the following statements **BEST** describes a key activity in data preparation for machine learning?

a) Data preparation comprises the identification and gathering of data from various sources, providing the algorithm with raw data to learn from

b) Feature engineering of the data is performed after an ML model has been trained to optimize its performance for real-world deployment

c) Data pre-processing includes augmentation and sampling, which either add or reduce the number of examples in the training data respectively

d) Exploratory data analysis (EDA) is a form of exploratory testing applied to the data preparation activities of acquisition, pre-processing and labelling

Select ONE answer.

Question #14 (1 Point)

Which of the following statements **BEST** contrasts the roles of training, validation, and test datasets in ML model development?

a) The training dataset is used to optimize hyperparameters, the validation dataset is used to tune predictions, and the test dataset is used to generate training data

b) The training dataset is used to create the model, the validation dataset is used to tune the model, and the test dataset evaluates its performance on unseen data

c) The training dataset is used for final model evaluation, the validation dataset ensures the model does not overfit, and the test dataset is used for tuning hyperparameters

d) The training dataset ensures the model generalizes well, the validation dataset is used to deploy the model, and the test dataset is used for initial evaluation

Select ONE answer.

Question #15 (2 Points)

Consider the following confusion matrix for an image classifier:

Confusion Matrix	Predicted Positive	Predicted Negative
Actual Positive	78	6
Actual Negative	22	14

Which of the following options represents the **CORRECT** formula for calculating the precision of the classifier?

a) (20/120) *100

b) (78/120) *100

c) (78/100) *100

d) (22/100) *100

Select ONE answer.

Question #16 (1 Point)

During the training of a deep neural network, the network produces an output, and the loss is calculated.

Which of the following **CORRECTLY** describes the next step in the training process?

a) The weights and biases across the network are adjusted to reduce the calculated loss
b) The same training data is passed through the network again to confirm the loss value
c) The activation functions are altered to different non-linear formulas to find a better fit
d) The network's hidden layers are reset with new random weight values

Select ONE answer.

Question #17 (1 Point)

Given the following examples of AI-based systems:

i. A system that learns from real-time data to improve its failure predictions and automatically updates maintenance schedules
ii. A spam filter for an email app, which identifies spam based on predefined rules
iii. A recommendation engine on a streaming service that updates its suggestions based on a user's changing viewing habits and preferences
iv. A personal assistant that learns continuously from its user
v. A rule-based system for medical diagnosis

Which of the following **BEST** describes the systems which can be considered to be locked AI-based systems?

a) ii and v
b) i, iii, and iv
c) i, and iii
d) ii, iv, and v

Select ONE answer.

Question #18 (1 Point)

Which of the following options BEST explains why a statistical approach is necessary when testing an AI-based system?

a) The system is typically large and complex, making test automation impractical without the use of statistical sampling
b) The system exhibits non-deterministic behavior, requiring large and representative test datasets to draw meaningful conclusions
c) A single test case is sufficient to determine if a model is well-calibrated, but only statistical methods can verify accuracy
d) The system is trained on real-world data and therefore does not require a separate test dataset; instead, it requires statistical validation

Select ONE answer.

Question #19 (1 Point)

Why is setting a 'seed' only a **LIMITED** solution for addressing the test oracle problem in AI systems?

a) It is a method that only functions for AI systems built from complete and detailed technical specifications

b) It ensures reproducibility for individual test runs but cannot alter the model's inherent probabilistic nature in production

c) Its success depends upon extensive consultation with domain experts to select the most appropriate seed value for the tests

d) It introduces a high degree of subjectivity into the model's behavior, which complicates the test evaluation process.

Select ONE answer.

Question #20 (1 Point)

Which of the following statements **BEST** describes a common approach to testing GenAI models?

a) GenAI models are tested by verifying that their outputs exactly match predefined expected results

b) It involves manipulating diverse inputs and parameters and then assessing the output's adherence to rules, because a direct input-to-output match is not feasible

c) Since GenAI models are probabilistic, formal testing is unnecessary because outputs will always vary and cannot be evaluated effectively

d) GenAI models are mainly tested through manual review, as automated testing does not apply to creative AI outputs

Select ONE answer.

Question #21 (2 Points)

Your organization plans to deploy a GenAI-powered legal document generator. During implementation planning, the security team argues for focusing red teaming efforts on preventing prompt injection attacks, while the compliance team wants to prioritize bias detection in legal advice generation. The development team suggests running red teaming after the system goes live to save time.

Which of the following is the **MOST** effective red teaming implementation strategy for this scenario?

a) Prioritize security vulnerabilities first, then address bias issues in a separate phase after deployment

b) Focus on bias detection since legal advice accuracy is more critical than security concerns

c) Deploy immediately and conduct red teaming reactively based on incidents and user feedback

d) Use attack scenarios covering both security and bias vulnerabilities before deployment

Select ONE answer.

Question #22 (1 Point)

An ML-based weather prediction system provides excellent results for most locations; however, it has been noticed that the predictions for areas in the UK with an altitude greater than 1250 meters are regularly inaccurate.

Which of the following test levels should have been performed **MORE** thoroughly?

a) System testing
b) Input data testing
c) Component integration testing
d) ML model testing

Select ONE answer.

Question #23 (1 Point)

Which of the following statements about applying risk-based testing to ML systems is **CORRECT**?

a) Security and usability risks can apply to any system, while risks associated with data bias and ML model performance are specific to ML systems
b) Conventional systems handle risks related to data bias, whereas ML systems focus on risks related to algorithmic bias
c) Risk management in conventional systems is a static approach, while in a self-learning system risk needs to be adjusted dynamically
d) In conventional systems, functional correctness is the primary risk factor, whereas functional adaptability is the primary risk factor for an ML system

Select ONE answer.

Question #24 (1 Point)

The team developing a new machine learning model has received a dataset from a new, unverified third-party vendor. They are uncertain about the origin of this dataset and concerned that the raw data may have been tampered with.

Which of the following test approaches is **MOST** suitable for addressing this specific risk?

a) Data provenance testing
b) Data representativeness testing
c) Feature testing
d) Dataset constraint testing

Select ONE answer.

Question #25 (1 Point)

A financial services company has developed an ML system for loan approval. A tester responsible for testing this system wants to determine if there is potential bias in the system related to factors such as gender and age attributes.

Which of the following approaches would be **MOST** suitable for detecting bias in the training data early?

a) Conducting static analysis of the model's source code to identify how its operations on the age and gender related attributes could lead to bias

b) Testing using a dataset that is representative and analyzing the predictions for statistically significant differences in outcomes across age and gender

c) Reviewing the overall ML workflow and data preparation processes to identify potential sources of bias introduction

d) Performing disparate impact analysis using counterfactuals based on gender, age or both

Select ONE answer.

Question #26 (1 Point)

What is a KEY difference in the test strategy for a data pipeline built for training versus one designed for an operational system?

a) Testing an operational pipeline would focus mainly on validating individual transformation scripts, whereas a training pipeline's testing would prioritize end-to-end system testing

b) A training pipeline would rely almost exclusively on fault injection and back-to-back tests, while an operational pipeline would be limited to unit and component integration testing

c) Configuration management reviews would be critical for operational pipelines but are considered unnecessary for the less formal nature of exploratory training pipelines

d) Testing a training pipeline would primarily focus on determining that data is handled correctly, while an operational pipeline's testing would emphasize high performance and reliability under load

Select ONE answer.

Question #27 (1 Point)

What is the purpose of creating a 'reference dataset' during data representativeness testing?

a) It serves as a universally applicable dataset derived from trusted industry benchmarks

b) It functions as a quantitative baseline for formally comparing the statistical properties of the training data

c) It is used to apply stratified sampling directly to the training data to verify that all subgroups are covered

d) It serves as the primary dataset for performing the final model validation and testing with high independence

Select ONE answer.

Question #28 (2 Points)

You are testing an ML dataset for a banking loan approval system. The dataset contains the following attributes:

- applicant_id (integer, unique)
- annual_income (decimal, in USD)
- loan_amount (decimal, in USD)
- credit_score (integer, 300-850)
- employment_years (integer, 0-50)
- monthly_payment (decimal, in USD)

The following business rules apply:

- Loan amount cannot exceed 5 times annual income
- Monthly payment = loan_amount / 324 (for 30-year loans)
- Credit scores must be within the standard range
- All monetary values must be positive

When applying dataset constraint testing to validate the 'monthly_payment' for 30-year loans, which type of constraint would be **MOST** appropriate?

a) Single-value range constraint

b) Multi-value count constraint

c) Comparison correlate constraint,

d) Multi-value duplicate constraint

Select ONE answer.

Question #29 (1 Point)

Which of the following **BEST** explains the role of multiple annotations in data label correctness testing?

a) Multiple annotations can be used to automate label verification to ensure consistency

b) Multiple annotations can point to label defects based on the comparison of annotations

c) Multiple annotations compare label distributions across datasets to detect potential defects

d) Multiple annotations point to defects in data points with high model loss

Select ONE answer.

Question #30 (1 Point)

Given the following test techniques and test types:

1. ML functional performance testing
2. Testing for bias
3. Adversarial testing
4. Drift testing

And the following risks:

A. The ML model might perform differently for different demographic groups
B. Slightly modified inputs to the ML model might cause quite different and unexpected responses
C. Predictions made by the ML model might be inaccurate in some cases
D. ML model accuracy might have significantly decreased since it was deployed

Which of the following options **BEST** matches the test techniques and test types with the risks?

a) 1A – 2D – 3C – 4B
b) 1C – 2A – 3B – 4D
c) 1B – 2A – 3C – 4D
d) 1A – 2C – 3D – 4B

Select ONE answer.

Question #31 (1 Point)

Which of the following statements about the documentation of ML models is **CORRECT**?

a) Information related to the speed of model prediction is a part of the documentation of an AI component
b) Interaction of AI and non-AI components is within the scope of the model documentation
c) Bias-related characteristics of input data are challenging to assess as the model documentation doesn't contain the source of the training data
d) Changes made by self-learning ML systems are fully documented at the time of the update to ensure up-to-date documentation

Select ONE answer.

Question #32 (1 Point)

Upon release, an ML model provided correct predictions fewer times than expected. Additional tests have been conducted, and the level of accuracy for these tests was measured at 83%.

Previously, the ML model had achieved an accuracy of 83% ± 4% at a confidence level of 94%.

Which of the following options is **MOST** likely to represent the current situation?

a) accuracy of 83% ± 2% at 94% confidence level

b) accuracy of 83% ± 4% at 92% confidence level

c) accuracy of 83% ± 6% at 94% confidence level

d) accuracy of 85% ± 4% at 94% confidence level

Select ONE answer.

Question #33 (1 Point)

Which of the following statements **BEST** summarizes adversarial testing of machine learning systems?

a) It is a form of black-box testing, ignoring knowledge of the internals of the machine learning system to create adversarial examples

b) It focuses on generating manual adversarial examples, without automation, to test the vulnerability of machine learning systems

c) It identifies model vulnerabilities through adversarial examples, which are minimally perturbed inputs that induce misclassification

d) It verifies the machine learning system's functionality by using previously working tests to avoid ML model failures during evaluation and tuning

Select ONE answer.

Question #34 (2 Points)

An AI-based mobile phone search system provides a list of phones that it believes are most suitable for the user based on its knowledge of the user's previous mobile phone usage and their specified preferences.

Metamorphic testing is being used with the following source test case:

Inputs	
Selected price range:	$200-$300
3D camera:	Don't care
Screen size:	mid to large
OS:	Android or iOS
Battery Life:	Don't care

Outputs
Recommended Phones: SnapHappy_X1 SnapHappy_M2 SnapHappy_M3 ClickNow_1000x ClickNow_1000xs

And this test data for two corresponding follow-up test cases:

Input T1	
Selected price range:	$200-$300
3D camera:	yes
Screen size:	mid to large
OS:	Android or iOS
Battery Life:	Don't care
Input T2	
Selected price range:	$200-$300
3D camera:	no
Screen size:	mid to large
OS:	Android or iOS
Battery Life:	Don't care

Which of the following options is **MOST** likely to be a valid list of recommended phones for the follow-up test cases?

a) T1: SnapHappy_X1, SnapHappy_M2
T2: ClickNow_1000x, ClickNow_1000xs

b) T1: SnapHappy_M2, SnapHappy_M3, ClickNow_1000xs
T2: SnapHappy_X1, ClickNow_1000x

c) T1: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3, ClickNow_1000x, ClickNow_1000xs
T2: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3

d) T1: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3, ClickNow_1000x, ClickNow_1000xs
T2: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3, ClickNow_1000x, ClickNow_1000xs

Select ONE answer.

Question #35 (1 Point)

An organization uses an ML model to predict customer churn but has no mechanism to get direct user feedback or track when customers leave. They want to test for drift by analyzing the data being fed into the live ML model.

Which test type would be **MOST** appropriate in this situation and why?

a) Dynamic drift testing, because it can infer the current ground truth by analyzing the input data's statistical properties

b) Static drift testing, because it identifies drift by detecting changes in the data distributions without requiring ground truth

c) Dynamic drift testing, because comparing ML model predictions to actual results is the most direct way to measure performance degradation

d) Static drift testing, because it compares the ML model's live performance metrics against a predefined acceptance threshold

Select ONE answer.

Question #36 (1 Point)

When testing a trained ML model, the development team found that the model was highly accurate when evaluated using validation data but performed poorly with independent test data.

Which of the following options is **MOST** likely to cause this situation?

a) Underfitting

b) Concept drift

c) Overfitting

d) Low acceptance criteria

Select ONE answer.

Question #37 (1 Point)

Which of the following statements **BEST** describes how A/B testing is used in the context of a machine learning system (MLS)?

a) A/B testing is primarily used to generate diverse test cases that cover all possible inputs for an MLS

b) A/B testing is used to verify that all components within an MLS interact correctly with each other

c) A/B testing determines if an updated version of an MLS performs better than the previous version

d) A/B testing focuses on analyzing the internal algorithm structure of an MLS to identify potential defects

Select ONE answer.

Question #38 (1 Point)

Which of the following descriptions of the creation of a pseudo-oracle used to support the back-to-back testing of an ML model is **MOST** likely to be **CORRECT**?

a) By varying the hyperparameters used for training the ML model that is being tested

b) By fine-tuning the ML model that is being tested

c) By supplementing the ML model that is being tested with retrieval-augmented generation

d) By using a different ML development framework than the ML model that is being tested

Select ONE answer.

Question #39 (1 Point)

Which **TWO** of the following scenarios describe ML development risks that can be **EFFECTIVELY** mitigated by performing ML functional performance testing?

a) A team notices that their ML development framework becomes slow and even unresponsive when processing large batches of data

b) A new version of a core library used by the ML development framework appears to be causing the ML model to produce unexpected and inaccurate results

c) After a new installation, the development team needs a quick test to determine if the ML development framework's essential services are running correctly

d) A single test result is difficult to reproduce during evaluation, showing slight variations in ML functional performance with each execution

e) A project leader needs to decide between two potential algorithms based on their suitability for the project's goals

Select TWO answers.

Question #40 (1 Point)

Which of the following statements **CORRECTLY** describes how shadow testing differs from canary testing in the deployment of ML models?

a) Shadow testing affects the responses from real users, while canary testing does not involve real users

b) Shadow testing mirrors traffic without impacting users, while canary testing provides responses from the new ML model

c) Canary testing is used for performance testing, while shadow testing is mainly used for component integration testing

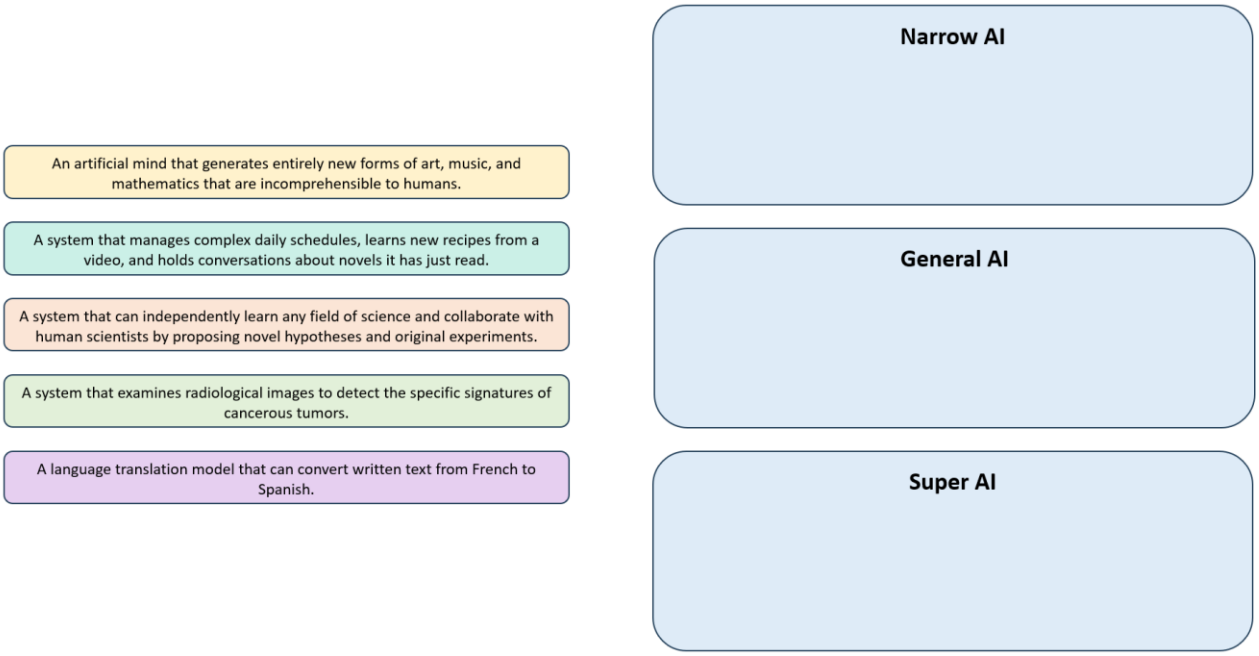
d) Canary testing compares ML models running offline, while shadow testing is based on the use of live user data

Select ONE answer.

AppendiDx: Additional Questions

Question #A1 (1 Point)

Given the following example AI systems on the left, and the three different forms of AI on the right:



Assign each example AI system to a form of AI. No form of AI can be left empty.

Question #A2 (1 Point)

Which of the following options **BEST** describes an advantage of using ML development frameworks when building and training machine learning models?

a) They provide support for specifying the ML model architecture design, such as the structure of a decision tree

b) They eliminate the need for data preprocessing by automatically optimizing the data before the training

c) They require the data scientists who use them to be advanced programmers with in-depth coding skills

d) They make model development highly efficient, but lock in any developed ML models to the ML development framework

Select ONE answer.

Question #A3 (1 Point)

An organization is developing an AI-based system for unmanned autonomous driving without considering the risks associated with the system.

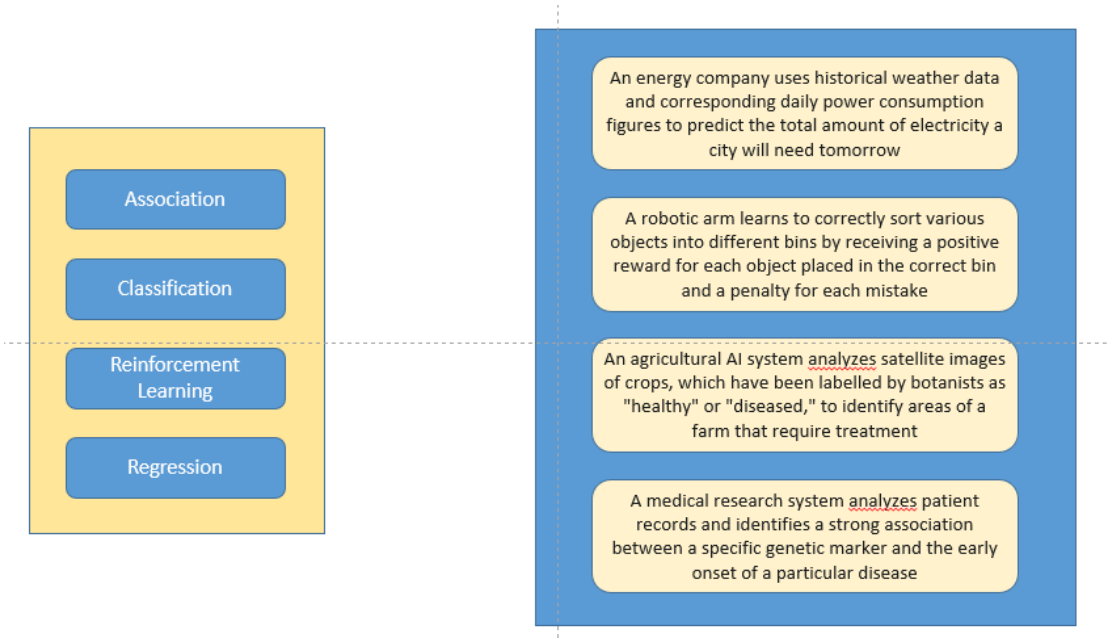
Which of the following standards is **MOST** likely being violated in this scenario, which could result in severe penalties?

a) ISO/IEC/IEEE TR 29119-11
b) OECD AI Principles
c) EU AI Act
d) ISO/IEC 25059

Select ONE answer.

Question #A4 (1 Point)

Given the following forms of machine learning, and example machine learning systems:



Drag to pair each form of machine learning with an example machine learning system.

Question #A5 (1 Point)

Given the following activities:

Order these activities in a logical sequence from earliest to latest in the ML workflow.

Question #A6 (1 Point)

What is a key reason that structural coverage alone is **NOT SUFFICIENT** as the basis for testing neural networks?

a) It tends to overestimate performance on tasks involving human feedback

b) It cannot account for differences in model size when comparing test results

c) It does not reveal whether the network relies on misleading or irrelevant features

d) It replaces the need to assess the usefulness of individual model outputs

Select ONE answer.