

Sample Exam – Questions

**Sample Exam set A
Version 1.0**

ISTQB® Certified Tester AI Testing Syllabus

Compatible with Syllabus version 1.0

International Software Testing Qualifications Board



Copyright Notice

Copyright Notice © International Software Testing Qualifications Board (hereinafter called ISTQB®).

ISTQB® is a registered trademark of the International Software Testing Qualifications Board.

All rights reserved.

The authors hereby transfer the copyright to the ISTQB®. The authors (as current copyright holders) and ISTQB® (as the future copyright holder) have agreed to the following conditions of use:

Extracts, for non-commercial use, from this document may be copied if the source is acknowledged.

Any Accredited Training Provider may use this sample exam in their training course if the authors and the ISTQB® are acknowledged as the source and copyright owners of the sample exam and provided that any advertisement of such a training course is done only after official Accreditation of the training materials has been received from an ISTQB®-recognized Member Board.

Any individual or group of individuals may use this sample exam in articles and books, if the authors and the ISTQB® are acknowledged as the source and copyright owners of the sample exam.

Any other use of this sample exam is prohibited without first obtaining the approval in writing of the ISTQB®.

Any ISTQB®-recognized Member Board may translate this sample exam provided they reproduce the abovementioned Copyright Notice in the translated version of the sample exam.

Document Responsibility

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

Acknowledgements

This document was produced by a core team from ISTQB®: Klaudia Dussa-Zieger, Werner Henschelchen, Vipul Koch, Qin Liu, Stuart Reid, Kyle Siemens, and Adam Leon Smith.

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

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

Revision History

Sample Exam – Questions Layout Template used: Version 2.5 Date: May 21st, 2021

Version	Date	Remarks
1.0	2021/10/01	Release for GA

Table of Contents

Copyright Notice	2
Document Responsibility	2
Acknowledgements	2
Revision History	3
Table of Contents	4
Introduction	6
Purpose of this document	6
Instructions	6
Questions	7
Question #1 (1 Point)	7
Question #2 (1 Point)	7
Question #3 (1 Point)	7
Question #4 (1 Point)	8
Question #5 (1 Point)	8
Question #6 (1 Point)	9
Question #7 (1 Point)	9
Question #8 (1 Point)	10
Question #9 (1 Point)	10
Question #10 (1 Point)	11
Question #11 (2 Points)	11
Question #12 (1 Point)	12
Question #13 (1 Point)	12
Question #14 (1 Point)	12
Question #15 (1 Point)	13
Question #16 (1 Point)	13
Question #17 (2 Points)	14
Question #18 (1 Point)	14
Question #19 (2 Points)	15
Question #20 (1 Point)	15
Question #21 (1 Point)	16
Question #22 (1 Point)	16
Question #23 (1 Point)	16
Question #24 (1 Point)	17
Question #25 (2 Points)	17
Question #26 (1 Point)	18
Question #27 (1 Point)	18
Question #28 (1 Point)	18
Question #29 (1 Point)	19
Question #30 (1 Point)	19
Question #31 (1 Point)	20
Question #32 (1 Point)	20
Question #33 (2 Points)	21
Question #34 (1 Point)	22
Question #35 (2 Point)	23
Question #36 (1 Point)	23
Question #37 (1 Point)	24
Question #38 (1 Point)	24
Question #39 (1 Point)	24
Question #40 (1 Point)	25
Appendix: Additional Questions	26
Question #1 (1 Point)	26
Question #2 (1 Point)	26

Question #3 (1 Point)	26
Question #4 (1 Point)	27
Question #5 (1 Point)	27
Question #6 (1 Point)	27
Question #7 (1 Point)	28
Question #8 (1 Point)	28
Question #9 (1 Point)	28
Question #10 (1 Point)	29
Question #11 (1 Point)	29
Question #12 (1 Point)	29
Question #13 (1 Point)	30
Question #14 (1 Point)	30
Question #15 (1 Point)	30
Question #16 (1 Point)	31
Question #17 (1 Point)	31
Question #18 (1 Point)	31
Question #19 (1 Point)	32
Question #20 (1 Point)	32
Question #21 (1 Point)	32
Question #22 (1 Point)	33
Question #23 (1 Point)	33
Question #24 (1 Point)	33
Question #25 (1 Point)	34
Question #26 (1 Point)	34
Question #27 (1 Point)	34

Introduction

Purpose of this document

The sample questions and answers and associated justifications in this sample exam set 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.

These questions cannot be used as-is in any official examination, but they should serve as guidance for question writers. Given the wide variety of formats and subjects, these sample questions should offer many ideas for the individual Member Boards on how to create good questions and appropriate answer sets for their examinations.

Instructions

In this document you may find:

- Questions¹, 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 provides the **BEST** example of the 'AI Effect'?

- a) People lose their jobs as AI-based systems perform their roles cheaper and better
- b) Competitive computer games lose popularity as AI-based systems always win
- c) Rule-based expert systems for medical diagnosis are no longer considered to be AI
- d) People believe AI will take over the world, as shown in films

Select ONE option.

Question #2 (1 Point)

Which of the following options is **NOT** a technology used to implement AI?

- a) Support vector machine
- b) Decision tree
- c) Evolutionary reasoning
- d) Bayesian optimization

Select ONE option.

Question #3 (1 Point)

Which of the following statements about the hardware used to implement AI-based systems is **MOST** likely to be **CORRECT**?

- a) The processors used to train a mobile recommendation system must be the same as the processors on the mobile phone
- b) Graphical processing units (GPUs) are a reasonable choice to implement an AI-based computer vision system
- c) Deep learning systems need to be trained, evaluated, and tested using AI-specific chips
- d) It is always best to choose processors with more bits to achieve sufficient accuracy for AI-based systems

Select ONE option.

Question #4 (1 Point)

There are a number of good quality pre-trained models available in the market and you want to use one of them for an image-based classifier. You have decided to ask the provider of the model about the data used for training the model and its format.

Which of the following statements is the **BEST** example of a risk that you are trying to mitigate by asking these questions?

- a) Bad classification accuracy of the pre-trained models
- b) Differences in the data used to train the model and the operational data
- c) Performance efficiency issues of the pre-trained model
- d) Lack of explainability of the pre-trained model compared to that of a model trained by you

Select ONE option.

Question #5 (1 Point)

Which of the following statements is **MOST** likely to be specifying a requirement for autonomy in an AI-based system?

- a) The system shall maintain a safe distance to other vehicles until the brake or accelerator is pushed by the driver
- b) The system shall learn the preferred style of response to emails by remotely monitoring the email traffic
- c) The system shall compare its predictions of house prices with actual selling prices to determine if it needs to be retrained
- d) It shall be possible to modify the system's behavior to work with different types of users in less than a day

Select ONE option.

Question #6 (1 Point)

Which of the following statements about bias in AI-based systems is **NOT** correct?

- a) Bias may be caused by users of a book recommendation system making choices that deliberately cause the system to make poor suggestions
- b) Bias may be caused in the employee age of death prediction system by collecting the training data from a dataset of patients who are all retired
- c) Bias may be caused in the creditworthiness system by using training data obtained from those who own and use a credit card
- d) Bias may be caused in the navigation system by using a route planning algorithm that is too complex to be explained to typical users

Select ONE option.

Question #7 (1 Point)

Which of the following is **MOST** likely to be an example of reward hacking?

- a) The programmer's assistant tool optimizes the code to provide reduced response times, while still ensuring that functional requirements are met
- b) An anesthetic supply device with a goal of keeping patients stable during surgery supplies too many doses and patients do not wake up as quickly as expected
- c) The third-party development organization paid their AI programmers based on the number of lines of code they write
- d) A type of AI used to play competitive computer games against humans that is focused on getting the highest score

Select ONE option.

Question #8 (1 Point)

Given the following attributes for an AI-based system (I-V):

- I. Probabilistic
- II. Explicable
- III. Unfair
- IV. Non-deterministic
- V. Deterministic

Which list of attributes below is likely to cause the **MOST** difficulties if the system is to be used as part of a safety-related system?

- a) I, IV
- b) II, IV
- c) II, III, V
- d) I, III, V

Select ONE option.

Question #9 (1 Point)

Which of the following statements BEST describes classification and regression as part of supervised learning?

- a) Regression is checking that the ML model test results do not change when the same test data is executed
- b) Classification is the grouping of unlabeled data into separate classes
- c) Classification is the labelling of the data for training the ML model
- d) Regression is predicting the number of classes that are output by the ML model

Select ONE option.

Question #10 (1 Point)

Which of the following options **BEST** describes an example of reinforcement learning?

- 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) The factory quality control system uses video cameras and audio analysis to identify manufactured items that are faulty based on monitoring a human quality control operative
- d) The software component test prediction system uses a range of quality measures to identify which components are likely to contain the most defects

Select ONE option.

Question #11 (2 Points)

You have been asked for your opinion on the ML approach to be used for a new system that is part of the traffic management for a SMART city. The idea is that the new system will control the traffic lights in the city to ensure traffic flows easily through and around the city.

Which of the following approaches do you expect **MOST** likely to succeed?

- a) Unsupervised learning that is based on identifying clusters around the city where the traffic density is higher than average
- b) A supervised learning regression solution based on thousands of journeys labelled with both journey length and duration
- c) Reinforcement learning that is based on a reward function that penalizes solutions that result in higher levels of traffic congestion
- d) A supervised learning classification solution that is based on drivers and passengers submitting their favorite routes for traversing the city

Select ONE option.

Question #12 (1 Point)

When performing testing of a trained model, an ML engineer found that the model was highly accurate when evaluated with validation data but that it 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) Poor acceptance criteria

Select ONE option.

Question #13 (1 Point)

Which of the following is an example of a challenge that is likely to be encountered in the course of developing and testing an ML solution?

- a) Data anonymization operations typically require knowledge of various ML algorithms
- b) The data used might be unstructured data
- c) A large percentage of the budget gets spent just in data preparation
- d) The data pipeline scalability is a challenge when training the model

Select ONE option.

Question #14 (1 Point)

The data scientist has complained that the model cannot be trained with one particular algorithm, although other algorithms work with the same training data.

Which of the following options is the **MOST** likely reason for this?

- a) Wrong data
- b) Missing data
- c) Badly labelled data
- d) Insufficient data

Select ONE option.

Question #15 (1 Point)

DataSure is a start-up with a product that promises to improve the quality of ML models. DataSure claim that this improvement comes from checking if the data has been labeled correctly.

Which of the following defects is **MOST** likely to have been prevented by using this product?

- a) The model will have security vulnerabilities
- b) The model will have poor accuracy
- c) The model will not fulfill its intended function
- d) The model will produce biased outputs

Select ONE option.

Question #16 (1 Point)

An ML engineer, upon finding insufficient training data, is rotating labeled images to create additional training data.

Which of the following approaches to labeling is being applied in this above example?

- a) Crowdsourcing
- b) Augmentation
- c) AI-based labeling
- d) Outsourcing

Select ONE option.

Question #17 (2 Points)

The confusion matrix for an image classifier is shown below.

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

Which of the following options represents the precision of the classifier?

- a) $20/120 * 100$
- b) $78/120 * 100$
- c) $78/100 * 100$
- d) $22/100 * 100$

Select ONE option.

Question #18 (1 Point)

ThermalSpace is a solution provider that helps thermal power plants to optimize their power output. Their solution is based on an ML model created using past data with clearly marked output. The model helps determine the amount of electricity to be generated at a given time of the day.

To determine the quality of the model using ML functional performance metrics, which of the following metrics is **MOST** likely to be used?

- a) R-squared
- b) Precision
- c) Recall
- d) False Positives

Select ONE option.

Question #19 (2 Points)

KnowYourPet is an app utilizing ML to determine whether a pet is hungry or not. It is understood that a dog is likely to be not hungry most of the time, as reflected in the training data. If the dog is mis-diagnosed as hungry then it may lead to overfeeding of the dog and this could lead to serious health issues.

Which of the following metrics would you choose for determining the suitability of the model under test?

- a) Accuracy
- b) Precision
- c) Recall
- d) F1-score

Select ONE option.

Question #20 (1 Point)

Which of the following options **BEST** describes a deep neural net?

- a) It is comprised of a hierarchical structure of neurons with the lowest (deepest) neurons making most of the decisions
- b) It is comprised of connected neurons where each neuron has an associated bias and each connection has an associated weight
- c) It is made up several layers with each layer (except input and output layers) connected to each other layer and errors are propagated backwards through the network
- d) It is made up of layers of neurons, each of which generates an activation value based on the other neurons in the same layer

Select ONE option.

Question #21 (1 Point)

Which of the following statements **CORRECTLY** describes a test coverage measure for neural networks?

- a) Value change coverage is based on individual neurons being seen to affect the overall output of the neural network
- b) Threshold coverage is based on neurons outputting an activation value greater than a pre-set value between zero and one
- c) Neuron coverage is a measure of the proportion of neurons that are activated at any time during the testing
- d) Sign change coverage measures the coverage of neurons that output both positive, negative and zero activation values

Select ONE option.

Question #22 (1 Point)

Which of the following requirements for an AI-based system is **MOST** likely to cause a significant challenge in testing?

- a) The system shall be more accurate than the system it is replacing
- b) The AI component in the system shall have 100% accuracy
- c) A human operator should be able to override the system in 1 second
- d) The system shall mimic the human emotions of a typical game player

Select ONE option.

Question #23 (1 Point)

Which of the following is a factor associated with the test data that can make the testing of AI-based systems difficult?

- a) Sourcing big data with high velocity
- b) Sourcing data from a single source
- c) Sourcing data separately from the data scientists
- d) Sourcing data from public websites

Select ONE option.

Question #24 (1 Point)

Why would the accuracy of human decisions be considered in testing as well as the accuracy of AI-based systems?

- a) Intuitive human decisions can be made faster than a corresponding AI-based system in some situations
- b) Unethical decisions can be made by humans as well as AI-based systems
- c) The accuracy of human decisions is not relevant to testing AI-based systems
- d) Human decisions may be of lower quality when they have been recommended by an AI-based system

Select ONE option.

Question #25 (2 Points)

An ML-based toll charging solution determines the type of incoming vehicles from the images captured by a camera. There are different types of cameras available and the solution provider claims to be able to use cameras of different resolutions. The images need to be in jpeg format with a size of 320X480 pixels for the purpose of training the model as well as for predicting the outcome. The model should be able to classify the vehicle types with certain desired high level of accuracy and should be tested against vulnerabilities. Each toll plaza will have its own complete system unconnected to any other system.

Which of the following types of testing are the **MOST** appropriate options for the tests you would choose for system testing?

- a) Testing for concept drift
- b) Adversarial testing
- c) Scalability testing
- d) Fairness testing
- e) Data pipeline testing

Select TWO options.

Question #26 (1 Point)

Which of the following statements **BEST** describes a testing challenge that specifically applies to a self-learning system?

- a) The system requires regular retraining and therefore requires regular testing
- b) The system is regularly released which means regression testing is required
- c) The system changes in such a way that tests that previously passed can fail
- d) The system requires a human operator, who is also required for testing

Select ONE option.

Question #27 (1 Point)

Which of the following is **NOT** likely to be required to test a system for bias?

- a) Involving selected users that are known to be biased
- b) Measuring how changes in test inputs change test outputs
- c) Observing how production outputs correlate to production inputs
- d) Obtaining additional data from other sources

Select ONE option.

Question #28 (1 Point)

Which of the following statements **BEST** describes how system complexity can create challenges when testing an AI-based system?

- a) Testing for bias may require data that the team does not have
- b) Manual generation of white-box tests can be difficult
- c) Determining whether a system is ethical can be subjective
- d) It can be difficult to find representative data to train a model

Select ONE option.

Question #29 (1 Point)

An AI-based system is being used by the health ministry to identify vulnerable groups of patients, who will be provided with support and advice to help prevent them suffering future illnesses to which they may be susceptible. The results will also be shared with other government agencies and medical insurance companies. The system is initially being trained on a large set of data collected by the health ministry from two surveys of 5,000 men over 50 years of age and 25,000 women over 30 years of age. The system will continue to identify vulnerable patients by gathering information from publicly available social media.

Which of the following attributes should be **MOST** carefully considered when specifying the objectives and acceptance criteria for the system?

- a) Adaptability
- b) Bias
- c) Explainability
- d) Flexibility
- e) Autonomy

Select TWO options.

Question #30 (1 Point)

An ML engineer is trying to find exploitable inputs and then use these inputs to retrain the models to make them immune to these inputs.

Which of the following options **BEST** describes the approach being used by the ML engineer?

- a) Validation
- b) Adversarial testing
- c) Data pipeline testing
- d) Scalability testing

Select ONE option.

Question #31 (1 Point)

A test manager has to select test techniques to be used for testing autonomous vehicle software. There are a large number of environmental conditions (>50) that need to be considered for seven vehicle functions.

Which of the following test techniques is MOST likely to be used when testing the variety of vehicle functions (VF) in different environmental conditions (EC)?

- a) A/B testing based on the VF and EC parameters
- b) Combination testing of all the parameters of VF and EC
- c) Pairwise testing of the relevant values of VF and EC
- d) Back-to-back testing of relevant VF and EC values

Select ONE option.

Question #32 (1 Point)

A test manager decides to have a non-AI system with similar functionality to the AI based system under test (SUT) built to support system testing.

Which of the following statements is most likely to be **CORRECT**?

- a) The test manager has chosen back-to-back testing because it helps solve the test oracle problem by using a pseudo-oracle
- b) The test manager has chosen A/B testing because it helps solve the test oracle problem by using a pseudo-oracle
- c) The test manager has chosen back-to-back testing because the non-functional requirements of the SUT can be verified against the pseudo-oracle
- d) The test manager has chosen A/B testing because the non-functional requirements of the SUT can be verified against the pseudo-oracle

Select ONE option.

Question #33 (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.

Given that metamorphic testing is being used with the following source test case:

<u>Inputs</u>		<u>Outputs</u>
Selected price range:	\$200-\$300	Recommended Phones:
3D camera:	Don't care	SnapHappy_X1
Screen size:	mid to large	SnapHappy_M2
OS:	Android or iOS	SnapHappy_M3
Battery Life:	Don't care	ClickNow_1000x
		ClickNow_1000xs

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

<u>Input T1</u>	
Selected price range:	\$200-\$300
3D camera:	yes
Screen size:	mid to large
OS:	Android or iOS
Battery Life:	Don't care
<u>Input T2</u>	
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 option.

Question #34 (1 Point)

System testing of an AI-based system is being planned. It has been suggested that exploratory testing is used in addition to scripted test techniques.

Which of the following scenarios is **MOST** likely to be an example of exploratory testing being performed?

- a) Training data is visualized using tools to look at various aspects of the data
- b) Tests written using equivalence partitioning during the previous test cycle are being run
- c) The Google 'ML test checklist' is being used
- d) ML functional performance metrics are being calculated

Select ONE option.

Question #35 (2 Point)

LAIgal systems has an AI-based product for extracting relevant favorable judgements similar to a given legal case. This product is used by judges in the courts. Details of the current case are provided, and the system produces relevant judgements. The system needs to be safe from malicious inputs. A similar open-source product exists and is available. Not having a suitable test oracle is a challenge when testing.

Which of the following test techniques should be selected to test the new version during system testing?

- a) A/B testing
- b) Back-to-back testing
- c) Adversarial testing
- d) State transition testing
- e) ML functional performance metrics calculation

Select TWO options.

Question #36 (1 Point)

Which one of the following statements is an example of a difference between a test environment for AI-based systems and a test environment for conventional systems?

- a) Test environments for AI-based systems may require some mechanism to determine how a particular decision is made
- b) Test environments for AI-based systems need simulators and virtual environments whereas conventional systems do not need these
- c) Test environments for AI-based systems need large amounts of data, whereas conventional systems do not need large amount of data
- d) GPUs are required for test environments for AI-based systems whereas conventional systems do not need these

Select ONE option.

Question #37 (1 Point)

In which of the following situations would AI be **MOST** useful when categorizing new defects?

- a) A small number of defects requires categorization on a new application
- b) A large number of defects is reported on a small application
- c) Minimal data is provided in typical defect reports
- d) A new development team needs to know the most appropriate developer to fix a defect

Select ONE option.

Question #38 (1 Point)

Which of the following is an AI tool **MOST** likely to use as the basis for generating functional test cases?

- a) A test charter
- b) A picture of the system as a flow chart
- c) Web server logs
- d) Crash reports

Select ONE option.

Question #39 (1 Point)

Which of the following options **CORRECTLY** states how an AI-based tool can perform optimization of regression test suites?

- a) By analyzing false positive test results
- b) By analyzing information from previous testing activities
- c) By using genetic algorithms to create new test cases
- d) By updating the expected results to counter concept drift

Select ONE option.

Question #40 (1 Point)

Which of the following options **CORRECTLY** states how an AI-based tool can perform defect prediction?

- a) Using natural language to ask developers where they predict defects will occur
- b) By analyzing the causes of defects raised on a similar code base
- c) By analyzing false positive defects
- d) Scanning code to identify defects using rules.

Select ONE option.

Appendix: Additional Questions

Question #1 (1 Point)

Which of the following statements about AI is **MOST** likely to be **CORRECT**?

- a) An autonomous robot that can act as a worker in a house, shop or office is an example of general AI
- b) A robot exhibiting similar skill levels as a human is considered to have achieved singularity
- c) AI-based systems that support a range of test management functions are considered to possess general AI
- d) An AI-based system that cannot access the internet is said to exhibit narrow AI

Select ONE option.

Question #2 (1 Point)

Which of the following statements is MOST likely to be describing a conventional system (as opposed to an AI-based system)?

- a) This system assigns customers into groups, based on their historical buying patterns
- b) This system controls the braking of the car dependent on its speed
- c) This system taught itself to recognize different words by listening to recordings
- d) This system detects anomalies from its experience of seeing anomalies in many X-rays

Select ONE option.

Question #3 (1 Point)

Which of the following options is **NOT** a framework used to develop AI-based software?

- a) scikit-learn
- b) CNTK
- c) MxNet
- d) EZPy-AI

Select ONE option.

Question #4 (1 Point)

Which of the following statements is **MOST** likely to be describing a system that includes the use of AI as a Service (AlaaS)?

- a) The image classifier identifies defects in the gyroscope casings produced by the company and was built using a transfer learning approach so that it is of high accuracy
- b) The underwater AI-based vehicle steering system uses a third-party obstacle avoidance component based on decision trees and Bayesian optimization
- c) The contract checker uses an exclusive algorithm for determining levels of legal liability, but the pricing part of the contract is separately checked by a generic contract pricing AI component
- d) The car rental pricing system is built using AI to support a demand-based algorithm and is hosted in the cloud and made available to all of the company's car rental offices

Select ONE option.

Question #5 (1 Point)

Which of the following options is the **MOST** likely use of a regulatory standard/regulation for an AI-based system?

- a) Use of ISO/PAS 21448 (SOTIF) for an unmanned autonomous submarine
- b) Use of GDPR for a bank loan decision-making system
- c) Use of ISO 26262 for a fully self-driving car
- d) Use of GDPR for a drone collision-avoidance system

Select ONE option.

Question #6 (1 Point)

Which of the following statements about flexibility and adaptability is **MOST** likely to be **CORRECT**?

- a) Adaptability is important in unsupervised learning as it allows the ML model to learn from data without labels
- b) Flexibility is important in supervised learning as it allows the ML model to recognize meaning even when data is poorly labelled
- c) Adaptability is important in reinforcement learning systems as such systems must adapt themselves to optimize their reward function
- d) Flexibility is important in self-learning systems as it allows them to adapt themselves to unexpected changes in their environment

Select ONE option.

Question #7 (1 Point)

Which of the following statements about the evolution of AI-based systems is **CORRECT**?

- a) Self-learning AI-based systems that continue to work in the same operational environment are not expected to change their behavior
- b) Side effects are not a concern for AI-based systems that change themselves to cope with changes in their environment
- c) AI-based systems must change themselves to cope with changes in system requirements during development
- d) Self-learning systems that physically interact with people, need to be managed to ensure system changes are not dangerous

Select ONE option.

Question #8 (1 Point)

Which of the following examples of an AI-based system is **LEAST** likely to require special attention with regard to ethical issues?

- a) A computer game that teaches children the benefits of democracy by challenging them to become president
- b) An application that uses data available on social media to provide a trustworthiness rating for job applicants
- c) An autonomous self-powered underwater marine mapping system to create a map of the seabed in international waters
- d) A mobile app that monitors each employee's daily exercise and rewards them with health insurance benefits

Select ONE option.

Question #9 (1 Point)

Which of the following statements about the transparency, interpretability and explainability for AI-based systems is **MOST** likely to be **CORRECT**?

- a) The search engine algorithm used for training students on search engine technology was selected as it was considered to be the most explainable
- b) The loan system was considered transparent as for each loan application it was clear to users how it decided whether to approve agree to the loan or not
- c) The doctors were happy with the level of interpretability of the rule-based oncology system as they could understand how the given rules were implemented in the system
- d) The drone operators were happy with the transparency of the control system as they felt that the system responded correctly to their instructions

Select ONE option.

Question #10 (1 Point)

Which of the following **BEST** describes the unsupervised approach to machine learning?

- a) Data and labels are analyzed to group them into clusters
- b) A system automatically learns by satisfying a fitness function
- c) A system teaches itself to meet objectives based on rewards
- d) Data are analyzed to identify patterns in the data

Select ONE option.

Question #11 (1 Point)

Given the following descriptions:

- I. Model performance is checked using validation data
- II. The origin of test data used to test the model is identified
- III. The tuned model is made ready for its target hardware
- IV. Test data are used to ensure the agreed ML functional performance criteria are met
- V. The model is created from source code
- VI. The critical data features are identified

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

- a) III - Build & Compile Model / II - Prepare the Data / IV - Evaluate the Model
- b) V - Build & Compile Model / II - Prepare the Data / I - Evaluate the Model
- c) V - Build & Compile Model / VI - Prepare the Data / IV - Evaluate the Model
- d) III - Build & Compile Model / VI - Prepare the Data / I - Evaluate the Model

Select ONE option.

Question #12 (1 Point)

Which of the following statements is **LEAST** likely to be used as a rationale for selecting an ML algorithm?

- a) The amount of memory available for training the translation system in the mobile device
- b) The maximum time allowed for retraining the embedded health monitoring system
- c) The number of measured characteristics used as the basis for a sports prediction system
- d) The number of expected clusters of customer types for a retail marketing system

Select ONE option.

Question #13 (1 Point)

Which of the following statements about the test dataset is **CORRECT**?

- a) The test dataset comes from a source totally different from the validation dataset
- b) The format of the test dataset is different from the format of the validation dataset
- c) The test dataset can be used as the validation dataset but not as the training dataset
- d) The test dataset should not be exposed to the model during the training process

Select ONE option.

Question #14 (1 Point)

Which of the following options would **MOST** likely be a reason for poor labeling of data?

- a) Insufficient data
- b) Synthetic data
- c) Translation errors
- d) Algorithm chosen for the ML model

Select ONE option.

Question #15 (1 Point)

An ML team asserts that the ML functional performance metrics based on validation data collected as part of training an ML model are sufficient for determining the quality of the system.

Which of the following statements is a valid reason to show that this may be **INCORRECT**?

- a) The ML functional performance metrics may not work well if the ground truth is not correct
- b) The ML functional performance metrics cannot be used for measuring quality as these are tool dependent
- c) Validation data is biased resulting in skewed functional performance measurements
- d) Data may need to be transformed prior to training the model, so the functional performance measurements do not reflect the quality of the model

Select ONE option.

Question #16 (1 Point)

Which of the following options regarding benchmark suites, **BEST** completes the following statement?

- a) ML benchmark suites help choose a particular model by indicating the time it takes to train
- b) ML benchmark suites help choose a particular model by indicating the time it takes to test
- c) ML benchmark suites help choose a particular model by indicating the time it takes to validate
- d) ML benchmark suites help choose a particular model by indicating the time it takes to deploy

Select ONE option.

Question #17 (1 Point)

Which of the following test levels provides the **BEST** choice for performing bias-related testing?

- a) Component testing
- b) Input data testing
- c) System testing
- d) Model testing

Select ONE option.

Question #18 (1 Point)

Which of the following statements about the documentation of AI components is **CORRECT**?

- a) Because non-functional requirements are not a part of the documentation of an AI component, non-functional testing cannot be performed
- b) White-box testing of the interaction of AI and non-AI components is not possible if the interfaces are a part of the documentation
- c) Checking for bias in the data is made possible by including the source of the data in the documentation
- d) Self-adapting AI systems require each change made by the system to be fully documented

Select ONE option.

Question #19 (1 Point)

An ecommerce application recommends products to the user on the basis of their purchase history and purchases being made by other people using the site, among other factors. As the tester, you have been asked to measure the current conversion rate of recommendations to compare with the original required conversion rate.

Which of the following is the **MOST** likely underlying reason for this request?

- a) AI effect
- b) Adversarial attacks
- c) Concept drift
- d) Lack of fairness

Select ONE option.

Question #20 (1 Point)

Which of the following options is **MOST** likely to be relevant when testing a system's autonomy?

- a) Testing over a sustained period of time
- b) Testing the accuracy of system predictions
- c) Testing how quickly the system can adapt
- d) Static analysis of training data

Select ONE option.

Question #21 (1 Point)

Which of the following statements demonstrates how non-deterministic systems can create challenges in testing?

- a) Non-deterministic systems produce a different result each time, normally preventing the generation of expected results
- b) Non-deterministic systems are difficult to test because they are not explainable, which hinders the investigation and fixing of defects
- c) A system that is given the same inputs and initial state may produce different outputs, so can require multiple test executions
- d) Non-deterministic systems are usually biased and require additional tests to allow this bias to be excluded from the results

Select ONE option.

Question #22 (1 Point)

When testing a deep neural network, which of the following characteristics can be evaluated without using dynamic testing?

- a) Explainability
- b) Transparency
- c) Automation bias
- d) Accuracy

Select ONE option.

Question #23 (1 Point)

Which of the following characteristics of an example AI-based system might cause a test oracle problem?

- a) It is not known where the training data was obtained
- b) The output of the system is a prediction for which the ground truth is unknown
- c) The system runs with no human intervention and is considered to be autonomous
- d) There is a lack of transparency into how the system was implemented

Select ONE option.

Question #24 (1 Point)

Which of the following statements is **CORRECT** with regards to A/B testing?

- a) A/B testing is also known as differential testing because two different programs are used for this type of testing
- b) A/B testing is mostly useful for testing simple ML models as it does not produce accurate results for complex models
- c) A/B testing requires multiple expected results from the same inputs to be compared to identify significant differences in tested models
- d) A/B testing is a good technique for writing test cases for various types of ML models, especially self-learning systems

Select ONE option.

Question #25 (1 Point)

Which of the following statements is **LEAST** likely to be an example of a benefit provided by virtual test environments in the testing of an AI-based system?

- a) A smart city application is tested in a virtual test environment to allow scenarios that very rarely happen (e.g., crowd control at a new year event) to be tested
- b) An AI-based money market trading system is tested in a virtual test environment as it would be impractical to test on real money markets due to the potential costs
- c) A smart fruit picker is tested in a virtual test environment to allow many picking scenarios to be run in an accelerated timeframe to check that fruit is not damaged
- d) An autonomous car is tested in a virtual test environment to allow potentially dangerous test scenarios to be run safely

Select ONE option.

Question #26 (1 Point)

It is possible to use an AI-assisted test automation tool to recognize objects through image processing, rather than using references to their location.

To which type of AI software engineering technology is this example **MOST** likely referring?

- a) Classification, Learning and Prediction
- b) Probabilistic Software Engineering
- c) Search-based Software Engineering
- d) Clustering

Select ONE option.

Question #27 (1 Point)

Which of the following statements is **CORRECT** with respect to visual testing?

- a) Visual testing makes use of object IDs to determine changes
- b) Visual testing uses images to do pixel-by-pixel comparisons
- c) Visual testing helps find overlapping user interface elements
- d) Visual testing fails when the screen layout changes

Select ONE option.

Sample Exam – Answers

**Sample Exam set A
Version 1.0**

ISTQB® Certified Tester AI Testing Syllabus

Compatible with Syllabus version 1.0

International Software Testing Qualifications Board



Copyright Notice

Copyright Notice © International Software Testing Qualifications Board (hereinafter called ISTQB®).

ISTQB® is a registered trademark of the International Software Testing Qualifications Board.

All rights reserved.

The authors hereby transfer the copyright to the ISTQB®. The authors (as current copyright holders) and ISTQB® (as the future copyright holder) have agreed to the following conditions of use:

Extracts, for non-commercial use, from this document may be copied if the source is acknowledged.

Any Accredited Training Provider may use this sample exam in their training course if the authors and the ISTQB® are acknowledged as the source and copyright owners of the sample exam and provided that any advertisement of such a training course is done only after official Accreditation of the training materials has been received from an ISTQB®-recognized Member Board.

Any individual or group of individuals may use this sample exam in articles and books, if the authors and the ISTQB® are acknowledged as the source and copyright owners of the sample exam.

Any other use of this sample exam is prohibited without first obtaining the approval in writing of the ISTQB®.

Any ISTQB®-recognized Member Board may translate this sample exam provided they reproduce the abovementioned Copyright Notice in the translated version of the sample exam.

Document Responsibility

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

Acknowledgements

This document was produced by a core team from the ISTQB®: Klaudia Dussa-Zieger, Werner Henschelchen, Vipul Koch, Qin Liu, Stuart Reid, Kyle Siemens, and Adam Leon Smith.

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

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

Revision History

Sample Exam – Answers Layout Template used: Version 2.5 Date: May 21st, 2021

Version	Date	Remarks
1.0	2021/10/01	Release for GA

Table of Contents

Copyright Notice	2
Document Responsibility	2
Acknowledgements	2
Revision History	3
Table of Contents	4
Introduction	6
Purpose of this document	6
Instructions	6
Answer Key	7
Answers	8
1	8
2	8
3	9
4	9
5	10
6	10
7	11
8	11
9	12
10	12
11	13
12	13
13	14
14	14
15	15
16	15
17	15
18	15
19	16
20	16
21	17
22	17
23	17
24	18
25	18
26	19
27	19
28	20
29	21
30	22
31	22
32	22
33	23
34	23
35	24
36	24
37	25
38	25
39	26
40	26
Appendix: Answers to Additional Questions	27
1	27

2	27
3	28
4	28
5	28
6	29
7	29
8	30
9	30
10	30
11	31
12	32
13	32
14	32
15	33
16	33
17	33
18	34
19	34
20	34
21	35
22	35
23	35
24	36
25	36
26	37
27	37

Introduction

Purpose of this document

The sample questions and answers and associated justifications in this sample exam set 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.

These questions cannot be used as-is in any official examination, but they should serve as guidance for question writers. Given the wide variety of formats and subjects, these sample questions should offer many ideas for the individual Member Boards on how to create good questions and appropriate answer sets for their examinations.

Instructions

In this document you may find:

- Answer Key table, including for each correct answer:
 - K-level, Learning Objective, and Point value
- Answer sets, including for all questions:
 - Correct answer
 - Justification for each response (answer) option
 - K-level, Learning Objective, and Point value
- Additional answer sets, including for all questions [does not apply to all sample exams]:
 - Correct answer
 - Justification for each response (answer) option
 - K-level, Learning Objective, and Point value
- *Questions are contained in a separate document*

Answer Key

Question Number (#)	Correct Answer	LO	K-Level	Points
1	c	AI-1.1.1	K2	1
2	c	AI-1.4.1	K1	1
3	b	AI-1.6.1	K2	1
4	b	AI-1.8.1	K2	1
5	a	AI-2.2.1	K2	1
6	d	AI-2.4.1	K2	1
7	b	AI-2.6.1	K2	1
8	a	AI-2.8.1	K1	1
9	b	AI-3.1.1	K2	1
10	a	AI-3.1.3	K2	1
11	c	AI-3.3.1	K3	2
12	c	AI-3.5.1	K2	1
13	c	AI-4.1.1	K2	1
14	d	AI-4.3.1	K2	1
15	b	AI-4.4.1	K2	1
16	b	AI-4.5.1	K1	1
17	c	AI-5.1.1	K3	2
18	a	AI-5.2.1	K2	1
19	b	AI-5.4.1	K4	2
20	b	AI-6.1.1	K2	1

Question Number (#)	Correct Answer	LO	K-Level	Points
21	b	AI-6.2.1	K2	1
22	d	AI-7.1.1	K2	1
23	a	AI-7.3.1	K1	1
24	d	AI-7.4.1	K2	1
25	b, e	AI-7.7.1	K4	2
26	c	AI-8.1.1	K2	1
27	a	AI-8.3.1	K2	1
28	b	AI-8.5.1	K2	1
29	b, c	AI-8.8.1	K4	2
30	b	AI-9.1.1	K2	1
31	c	AI-9.2.1	K2	1
32	a	AI-9.3.1	K2	1
33	b	AI-9.5.1	K3	2
34	a	AI-9.6.1	K2	1
35	b, c	AI-9.7.1	K4	2
36	a	AI-10.1.1	K2	1
37	b	AI-11.2.1	K2	1
38	c	AI-11.3.1	K2	1
39	b	AI-11.4.1	K2	1
40	b	AI-11.5.1	K2	1

Answers

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
1	c	<p>a) Is not correct. People in many occupations may lose their jobs to AI-based systems, but this is simply progress, not the 'AI Effect'.</p> <p>b) Is not correct. For some computer games, AI-based systems can outplay humans, but there is little evidence of a drop-off in the popularity of such games.</p> <p>c) Is correct. The 'AI Effect' is defined as the change in the definition of AI as technology advances. Rule-based systems for medical diagnosis were popular examples of AI in the 1970s and 1980s but are often not considered AI today.</p> <p>d) Is not correct. The gullibility of cinema goers believing killer robots will take over the world is not the 'AI Effect'.</p>	AI-1.1.1	K2	1
2	c	<p>a) Is not correct. Support vector machines are a form of machine learning.</p> <p>b) Is not correct. Decision trees are a form of machine learning.</p> <p>c) Is correct. There is no such AI technology as evolutionary reasoning. Such a term is occasionally used in discussions related to biological evolution-based rules were considered AI a few decades ago.</p> <p>d) Is not correct. Bayesian optimization is a form of machine learning.</p>	AI-1.4.1	K1	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
3	b	<p>a) Is not correct. The two activities of training a ML model and inference from that model are quite different so there is normally no reason that they should be performed on the same processors.</p> <p>b) Is correct. GPUs are designed for the parallel processing of images using thousands of cores, which is close to what is required for an AI-based computer vision system that would most likely be implemented as a neural network.</p> <p>c) Is not correct. It is still possible to train, evaluate and test a simple deep-learning system on a PC with limited GPU support – so specific chips for AI are not needed, but they would be far faster.</p> <p>d) Is not correct. Many AI-based systems are not focused on exact calculations, but rather on probabilistic determinations and so the accuracy of processors with many bits is often unnecessary.</p>	AI-1.6.1	K2	1
4	b	<p>a) Is not correct. The question mentions the good quality of the pre-trained model, so this risk should be negligible.</p> <p>b) Is correct. The data used to train the model should be similar to the data used for making the predictions.</p> <p>c) Is not correct. Performance does not appear to be an issue in this situation.</p> <p>d) Is not correct. Explainability does not appear to be an issue in this situation, nor can it be achieved by looking at the training data and its format.</p>	AI-1.8.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
5	a	<ul style="list-style-type: none"> a) Is correct. This requirement defines the human interventions that define the end of the system working autonomously. b) Is not correct. This requirement is specifying a required function for how the system shall perform self-learning. c) Is not correct. This requirement is specifying how the system will manage concept drift, in this case most likely caused by the house market changing. d) Is not correct. This is specifying an adaptability requirement – the maximum time it should take to make a change to the system. 	AI-2.2.1	K2	1
6	d	<ul style="list-style-type: none"> a) Is not correct. Bias can be caused by users deliberately poisoning the self-learning of an AI-based system. b) Is not correct. Bias can be caused when the training data does not correctly match those who the system will be applied to. For instance, employees will typically be younger than retired patients. c) Is not correct. Bias can be caused when the training data does not correctly match those who the system will be applied to. For instance, most people using credit cards are already considered creditworthy, which is a typical example of sample bias. d) Is correct. If the algorithm cannot be explained, then it lacks explainability, but that does not mean it is biased nor unbiased. 	AI-2.4.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
7	b	<p>a) Is not correct. It appears that the tool is achieving its two goals with there being no detrimental effects, so this is unlikely to be 'reward hacking'.</p> <p>b) Is correct. This could be 'reward hacking' if the system achieves one goal to the detriment of others, in this case the need for patients to wake up.</p> <p>c) Is not correct. Reward hacking is not a form of paying AI developers.</p> <p>d) Is not correct. Some game-playing AI-based systems are driven by a reward function, but this is not known as 'reward hacking'.</p>	AI-2.6.1	K2	1
8	a	<p>Considering the given attributes:</p> <ul style="list-style-type: none"> I. Probabilistic – a definite problem for safety-related systems as this causes non-determinism II. Explicable – normally needed for safety-related systems III. Unfair – not ideal, but sometimes unavoidable – and not a special problem for safety-related systems IV. Non-deterministic - a definite problem for safety-related systems V. Deterministic - normally needed for safety-related systems <p>Therefore, I and IV are the attributes that are MOST problematic for safety-related systems.</p> <p>a) Is correct.</p> <p>b) Is not correct</p> <p>c) Is not correct</p> <p>d) Is not correct</p>	AI-2.8.1	K1	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
9	b	<ul style="list-style-type: none"> a) Is not correct. Regression in the context of supervised learning is generally when the ML model outputs a numeric result. b) Is correct. Classification is when input data to a ML model is classified into one of a few predefined classes. c) Is not correct. Training data needs to be labelled for training in supervised learning, but this activity is not known as classification. It is simply labelling. d) Is not correct. Regression is when the output from the ML model is numeric, but the output is not a number of classes. 	AI-3.1.1	K2	1
10	a	<ul style="list-style-type: none"> a) Is correct. The amount spent can be considered the reward function for this system, with the system changing its behavior to increase the amount spent. b) Is not correct. The app is using text in what can be considered a source language and a 'correct' translation of this source. Therefore, it is relying on a form of supervised learning with no reward function mentioned. c) Is not correct. The system is using the human quality control operative as a form of 'gold' standard and so is relying on a form of supervised learning. d) Is not correct. There is no suggestion that any reward function is used, instead it is most likely that the prediction system bases its determination of defects on past experience. Therefore, it is probably also relying on a supervised learning system. 	AI-3.1.3	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
11	c	<p>a) Is not correct. It should be possible for the unsupervised learning system to identify areas that are congested, but this alone will not provide the solution.</p> <p>b) Is not correct. A regression solution is unlikely to provide us with what we want as the predicted speed of individual journeys will not provide an overall solution to citywide congestion.</p> <p>c) Is correct. A continually improving reinforcement learning system with a reward function based on lower levels of congestion as a measure of success is valid for this type of system.</p> <p>d) Is not correct. This solution is dependent on volunteers submitting subjective opinions that will most likely result in a solution that changes back and forth as the system adopts favorite routes that then become congested.</p>	AI-3.3.1	K3	2
12	c	<p>a) Is not correct. The model performs well on validation data, so it is not a case of underfitting.</p> <p>b) Is not correct. Concept drift refers to changes after the model training and validation stage.</p> <p>c) Is correct. The bad performance on test data and good on validation data suggests overfitting.</p> <p>d) Is not correct. Poor acceptance criteria should be consistent with different sets of data, so are unlikely to lead to a difference between the test results with validation data and independent test data.</p>	AI-3.5.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
13	c	<ul style="list-style-type: none"> a) Is not correct. Data anonymization operations do not require knowledge of ML algorithms. b) Is not correct. Unstructured data is not a challenge. Images, audio, free-flowing text are all examples of unstructured data. c) Is correct. Up to 36% of ML workflow effort may be spent in data preparation. d) Is not correct. Scalability typically is a requirement at deployment, rather than when training. 	AI-4.1.1	K2	1
14	d	<ul style="list-style-type: none"> a) Is not correct. Since models based on some learning algorithms can be trained with the data but not one particular algorithm, it indicates that the data is correct. b) Is not correct. Since models based on some learning algorithms can be trained with the data but not one particular algorithm, it indicates that there is no missing data. c) Is not correct. Since models based on some learning algorithms can be trained with the data but not one particular algorithm, it indicates that the data is correctly labelled. d) Is correct. Since models based on some learning algorithms can be trained with the data. However, if it does not work for one particular algorithm, it is MOST likely to be that the quantity of the data that is not sufficient for that particular algorithm. 	AI-4.3.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
15	b	<ul style="list-style-type: none"> a) Is not correct. Data privacy and security issues are not being handled. Hence the product is not going to prevent security issues. b) Is correct. Misabeled data results in reduced accuracy of the ML model. c) Is not correct. A model not being fit for purpose arises from is not correct or unfair data, not mislabeled data. d) Is not correct. A biased model results from incomplete data, unbalanced data, unfair data, data lacking diversity, or duplicate data, rather than from mislabeled data. 	AI-4.4.1	K2	1
16	b	<ul style="list-style-type: none"> a) Is not correct. Crowdsourcing is when you use a large number of people to provide some work. In this case only one person is performing the task. b) Is correct. Augmentation is being performed here by transforming existing labelled data. c) Is not correct. AI is not being used for labeling of the data. d) Is not correct. The ML engineer has not outsourced the task to a third party. 	AI-4.5.1	K1	1
17	c	<ul style="list-style-type: none"> a) Is not correct. See option c for the correct formula and calculation. b) Is not correct. See option c for the correct formula and calculation. c) Is correct. The formula for Precision = $TP / (TP + FP) * 100 = 78 / (78 + 22) = 78 / 100 * 100$ d) Is not correct. See option c for the correct formula and calculation. 	AI-5.1.1	K3	2
18	a	<ul style="list-style-type: none"> a) Is correct. It is a supervised regression problem because the model outputs a continuous value, the amount of electricity to be generated, which uses the R-Squared or MSE/RMSE metric. b) Is not correct. This is a metric for classification. c) Is not correct. This is a metric for classification. d) Is not correct. This is a metric for classification. 	AI-5.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
19	d	<p>a) Is not correct. Accuracy is not useful when there is an imbalance in the expected classes and the not hungry class dominates in this case.</p> <p>b) Is correct. Precision should be used because the cost of false-positives (overfeeding the dog) is high (serious health issues).</p> <p>c) Is not correct. Recall is useful when the positives should not be missed. In this case, precision is also important (see b) and hence recall alone is not very useful. F1-score is a better choice.</p> <p>d) Is not correct. F1-score is useful when there is an imbalance in the expected classes and when precision and recall are similarly important, but in this case precision appears to be far more important than recall.</p>	AI-5.4.1	K4	2
20	b	<p>a) Is not correct. A neural network does not have a hierarchical structure.</p> <p>b) Is correct. As with the human brain, an artificial neural network is comprised of connected neurons. To perform its calculation of an activation value, each neuron is assigned a bias and each connection is assigned a weight.</p> <p>c) Is not correct. A neural network is made up of several layers and errors are propagated backwards through the network, but the layers of a neural network are only connected to the next layers (not each other layer).</p> <p>d) Is not correct. A neural network is made up of layers of neurons, but the activation value is based on the neurons in the preceding layer (not the same layer).</p>	AI-6.1.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
21	b	<p>a) Is not correct. Value change coverage is a measure of the proportion of neurons activated where their activation values differ by more than a preset change amount. It is not concerned with the overall output of the neural network.</p> <p>b) Is correct. Threshold coverage measures the proportion of neurons activated during testing with a value greater than a preset threshold value.</p> <p>c) Is not correct. All neurons are potentially 'activated' each time a neural network is 'run', however the values output by the neurons change, which is what is measured by neuron coverage (coverage achieved by a value greater than zero).</p> <p>d) Is not correct. Sign change coverage is a measure of the proportion of neurons activated with both positive and negative activation values, but not zero activation values.</p>	AI-6.2.1	K2	1
22	d	<p>a) Is not correct. This is a specific requirement with a test oracle, so should not usually cause a testing challenge.</p> <p>b) Is not correct. This may be a difficult requirement to achieve but should not create a testing challenge.</p> <p>c) Is not correct. This is a testable requirement.</p> <p>d) Is correct. This requirement is extremely complex to test without defining all human emotions and how the system might mimic them.</p>	AI-7.1.1	K2	1
23	a	<p>a) Is correct. Sourcing data for AI systems that use large quantities of high-velocity data can be difficult.</p> <p>b) Is not correct. Sourcing consistent data from multiple sources can be difficult.</p> <p>c) Is not correct. Sourcing data separately is good practice as it prevents common failures with the data scientists.</p> <p>d) Is not correct. Sourcing data from public websites is straightforward.</p>	AI-7.3.1	K1	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
24	d	<ul style="list-style-type: none"> a) Is not correct. Speed of decision making is not related to accuracy. b) Is not correct. The ethical choices made by humans are not related to testing AI-based systems. c) Is not correct. The accuracy of human decisions is relevant as systems may make recommendations that humans approve or review. d) Is correct. Human decisions supported by recommendations by AI-based systems may be of lower quality than human decisions without recommendations from a system, and this should be considered in testing. 	AI-7.4.1	K2	1
25	b, e	<ul style="list-style-type: none"> a) Is not correct. Concept drift is tested after deployment. b) Is correct. Adversarial testing is important because the requirements state that the system should not be tested against vulnerabilities. c) Is not correct. Scalability testing has not been mentioned as one of the requirements. These are independent systems connected and are not connected to any other systems. d) Is not correct. Fairness is using positively biased data for training. Since there is no case of positive discrimination here, fairness testing is not relevant. e) Is correct. Data pipeline testing is required because the images can come in various formats and resolutions. For the model to be trained all images should have same format hence this testing is important. 	AI-7.7.1	K4	2

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
26	c	<ul style="list-style-type: none"> a) Is not correct. A system that requires regular retraining cannot be described as self-learning. b) Is not correct. A system that needs to be regularly released cannot be described as self-learning as it requires frequent releases to adapt to change. c) Is correct. Tests on a system that makes changes to itself may start to fail, even if they previously passed. d) Is not correct. A system that requires a human operator is unlikely to be self-learning. 	AI-8.1.1	K2	1
27	a	<ul style="list-style-type: none"> a) Is correct. Users that exhibit bias are not required to test a system for bias because they do not help to determine whether the behavior of the system is biased. b) Is not correct. Measuring how test inputs change test outputs is important when testing for bias as it can show how the system is biased towards or against particular inputs. c) Is not correct. Measuring how production inputs change production outputs is important when testing for bias because different results might be seen in production. d) Is not correct. Obtaining external data sources can be essential when testing for bias in case the bias is based on “hidden” variables. 	AI-8.3.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
28	b	<p>a) Is not correct. Bias does not usually relate to system complexity.</p> <p>b) Is correct. Understanding how the system works and creating enough tests to achieve effective coverage are challenges caused by the complexity of AI-based systems.</p> <p>c) Is not correct. Ethics is not usually related to AI-based system complexity.</p> <p>d) Is not correct. Difficulty finding representative data to train a model is not related to testing or AI-based system complexity.</p>	AI-8.5.1	K2	1

29	b, c	<p>Considering each of the attributes:</p> <ul style="list-style-type: none"> a) Adaptability - is the ability of the system to be modified (normally in order to continue to meet functional and non-functional requirements). There is no reason to believe the operational environment for the system will change much and so no reason to believe that the system will have to be changed. b) Bias – the data being used for training is biased towards women (25,000 vs 5,000) and towards specific age groups, therefore bias needs to be carefully considered. c) Explainability – the results may affect the identified vulnerable patients both medically and financially. They should be able to see why they have been labelled as susceptible so they can ensure they have been correctly chosen and as part of explainability requirements related to data privacy. d) Flexibility - is the ability of a system to change its behavior, but there is no reason at this point to think this system will need to be used outside the initial specification and so it is not expected there will be a need to change its behavior. e) Autonomy - is the ability of the system to work for sustained periods without human intervention. There is no reason to think that the system will have to work for sustained periods without intervention. <p>Thus, bias and explainability are the two most important attributes for the given scenario.</p> <ul style="list-style-type: none"> a) Is not correct. b) Is correct. c) Is correct. d) Is not correct. e) Is not correct. 	AI-8.8.1	K4	2
----	------	--	----------	----	---

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
30	b	<ul style="list-style-type: none"> a) Is not correct. This is not validation as the exploitable inputs are being found and retraining is being done using those inputs. b) Is correct. This is an example of adversarial testing. c) Is not correct. There are not any data pipelines that are being tested in this situation. d) Is not correct. No scalability tests are being performed in this example. 	AI-9.1.1	K2	1
31	c	<ul style="list-style-type: none"> a) Is not correct. A/B testing is not useful for combinatorial testing. b) Is not correct. All combinations would be almost impossible to do in practice, resulting from the near infinite number of potential combinations. c) Is correct. Pairwise testing is best suited to reduce the number of combinations without sacrificing defect detection too much. d) Is not correct. Back-to-back testing is not useful for combinatorial testing. 	AI-9.2.1	K2	1
32	a	<ul style="list-style-type: none"> a) Is correct. It is an example of back-to-back testing where the non-AI system is used as a pseudo-oracle. b) Is not correct. With A/B testing, we use a variant of the SUT to compare with the SUT. c) Is not correct. The resources and non-functional characteristics of the pseudo-oracle and the SUT are likely to be different, hence the alternate system cannot be used for non-functional testing. d) Is not correct. It is an example of back-to-back testing and also the resources and non-functional characteristics of the pseudo-oracle and the SUT are likely to be different. 	AI-9.3.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
33	b	<p>Follow-up test case T1 differs from the source test case by the change in requirements for a 3D camera; it is now more specific. A 3D camera must be included. So, that means the follow-up expected results can only include the original test results at most (the previously recommended phones with a 3D camera).</p> <p>Follow-up test case T2 also differs from the source test case by the change in requirements for a 3D camera; it is also more specific. No 3D camera should be included. So, that means the follow-up expected results can only include the original test results at most (the previously recommended phones without a 3D camera).</p> <p>As T1 lists phones <u>with</u> a 3D camera, the remaining phones from the source test case must be those with <u>no</u> 3D camera – and so they should be in T2.</p> <p>Therefore, T1 and T2 combined should contain all the cameras from the source test case, but with no overlap between the two.</p> <p>a) Is not correct. b) Is correct. c) Is not correct. d) Is not correct.</p>	AI-9.5.1	K3	2
34	a	<p>a) Is correct. This is Exploratory Data Analysis which is an exploratory method.</p> <p>b) Is not correct. This is scripted testing.</p> <p>c) Is not correct. This is checklist-based testing.</p> <p>d) Is not correct. Calculating ML functional performance metrics is not exploratory testing.</p>	AI-9.6.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
35	b, c	<p>a) Is not correct. A/B testing is most useful when comparing two variants for the purpose of deciding if the new variant is an improvement over the older variant.</p> <p>b) Is correct. Back-to-back testing uses a similar product as a pseudo-oracle for testing.</p> <p>c) Is correct. Adversarial testing is important here as it being used for very important purpose and adversarial data can cause harm.</p> <p>d) Is not correct. While state transition testing might be useful, nothing in the scenario suggests it; therefore, it is not the most relevant technique.</p> <p>e) Is not correct. This testing is appropriate at the model testing stage for classification problems. It is not appropriate at the system testing stage for non-classification problems.</p>	AI-9.7.1	K4	2
36	a	<p>a) Is correct. Explainability mechanism may need to be provided for AI environments.</p> <p>b) Is not correct. Simulators and virtual environments are often required for conventional systems.</p> <p>c) Is not correct. Large amount of data may be required for conventional systems, as well.</p> <p>d) Is not correct. GPUs may be required for many other systems as well, for example, games.</p>	AI-10.1.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
37	b	<p>a) Is not correct. Where a small number of defects requires categorization and there is no historical data, AI would not have training data to be used. Genetic algorithms and neural networks can be used for test generation, and even be combined. Clustering produces results applicable to test generation.</p> <p>b) Is correct. Where a large number of defects is reported on a small application there is most likely to be benefit and opportunity to identify duplicates.</p> <p>c) Is not correct. Where minimal data is provided in the defect reports, the usefulness of the tool will be lower, as less data will be available to the algorithm.</p> <p>d) Is not correct. For AI to recommend developers to fix defects it would need to be based on historical data. However, because a new development team is taking over, any recommendations would be inaccurate until historical data is available.</p>	AI-11.2.1	K2	1
38	c	<p>a) Is not correct. A test charter provides a focus for exploratory testing, and it rarely leads to the generation of test cases, even if an AI-based tool could interpret it.</p> <p>b) Is not correct. A flow chart could be used to generate tests, but it needs to be machine readable, rather than simply a picture.</p> <p>c) Is correct. Web server logs may reflect production use of the system and provide a way for AI to generate tests.</p> <p>d) Is not correct. Crash reports are unlikely to be used as they would describe unexpected failures rather than the functions performed by the application.</p>	AI-11.3.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
39	b	<p>a) Is not correct. The goal of regression testing optimization is to reduce the size, prioritize or augment a test suite, not to reduce false positives.</p> <p>b) Is correct. Optimization of regression test suites is performed by analyzing information on previous test executions.</p> <p>c) Is not correct. Per section 11.4 of the syllabus, regression test optimization is typically performed using previous test execution data. Using genetic algorithms to create new tests is unlikely to achieve the goal of optimizing the regression test suite.</p> <p>d) Is not correct. It is important to consider regression testing and concept drift together, however per section 11.4 concept drift is not related to regression test optimization using AI.</p>	AI-11.4.1	K2	1
40	b	<p>a) Is not correct. While natural language processing is an AI application, it is not used for defect prediction.</p> <p>b) Is correct. Defect prediction is performed by looking for correlations between code/process/people measures and defects on the same or a similar code base.</p> <p>c) Is not correct. The goal of defect prediction is not to identify defects with a false positive result. To analyze them would have little value.</p> <p>d) is not correct. Defect prediction does not involve scanning of code using rules. This is static analysis.</p>	AI-11.5.1	K2	1

Appendix: Answers to Additional Questions

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
1	a	<p>a) Is correct. General AI exhibits intelligent behavior comparable to a human and an autonomous robot that has a wide enough range of skills to perform as a worker in quite different environments is likely to be general AI.</p> <p>b) Is not correct. Singularity is that point when AI has surpassed humans – not when they are at similar levels.</p> <p>c) Is not correct. Performing test management would be considered by most as a single specialized set of tasks and would be classed as narrow AI.</p> <p>d) Is not correct. Narrow AI is being able to perform a single specialized task; it is not relevant whether the AI has access to the internet or not.</p>	AI-1.2.1	K2	1
2	b	<p>a) is not correct. This system appears to be describing clustering of customers as would be performed by an unsupervised learning system; therefore, it is likely to be AI-based.</p> <p>b) Is correct. There is no evidence that this system implements AI.</p> <p>c) Is not correct. This system appears to have programmed itself by listening to recordings; therefore, it is likely to be AI-based.</p> <p>d) Is not correct. This system appears to be describing a form of supervised learning based on anomalies in X-rays; therefore, it is likely to be AI-based.</p>	AI-1.3.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
3	d	<ul style="list-style-type: none"> a) Is not correct. This is a free software machine learning library for the Python programming language. b) Is not correct. This is the Microsoft Cognitive Toolkit (CNTK), an open-source deep-learning toolkit. c) Is not correct. This is a deep-learning open-source framework used by Amazon for AWS. d) Is correct. There is currently no AI development framework with this name 	AI-1.5.1	K1	1
4	c	<ul style="list-style-type: none"> a) Is not correct. The image classifier seems to be built for a highly specific problem area and is unlikely to be made widely available as AlaaS. b) Is not correct. Although the obstacle avoidance system is provided by a third party, it is highly likely to be embedded within the vehicle and it also appears to be a specialist component, so it is unlikely to be AlaaS. c) Is correct. The exclusive algorithm cannot be AlaaS, however the pricing part is generic and could well be provided as AlaaS. d) Is not correct. The described situation suggests that the system is made available across the web, but only internally to their own car rental offices, so it does not appear to be AlaaS. 	AI-1.7.1	K2	1
5	b	<ul style="list-style-type: none"> a) Is not correct. SOTIF is for road vehicles, not submarines. b) Is correct. A bank loan decision-making system will work with personal data and this is covered by GDPR. c) Is not correct. A fully self-driving car is likely to include non-deterministic systems, which are not allowed by ISO 26262. d) Is not correct. Drone collision systems are unlikely to include any personal data, which is the focus of GDPR 	AI-1.9.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
6	d	<p>a) Is not correct. In unsupervised learning the system is expected to learn from unlabeled data and adaptability as a characteristic is associated with a system being changed rather than changing itself.</p> <p>b) Is not correct. Supervised learning is dependent on labelled data, but flexibility is not a characteristic associated with reading poorly labelled data.</p> <p>c) Is not correct. Reinforcement learning systems are expected to optimize the reward function, and adaptability as a characteristic is associated with a system that can be changed rather than a system that changes itself.</p> <p>d) Is correct. Flexibility is a characteristic associated with the ability of systems to be used in contexts outside the original requirements.</p>	AI-2.1.1	K2	1
7	d	<p>a) Is not correct. Self-learning systems that work in the same operational environment could still optimize themselves and so change their behavior.</p> <p>b) Is not correct. AI-based systems that change themselves to adapt to changes in their environment could still cause negative side-effects.</p> <p>c) Is not correct. Evolution as a characteristic is not about system development, but about how an AI-based system changes after deployment in its operational environment.</p> <p>d) Is correct. If the self-learning system physically interacts with people, then any changes it makes to itself could potentially harm people.</p>	AI-2.3.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
8	c	<p>a) Is not correct. This game is aimed at children, a vulnerable group, and may give them an unfavorable view of democracy.</p> <p>b) Is not correct. The application may pick up private and/or biased information from social media, which adversely affects the job prospects of applicants.</p> <p>c) Is correct. The ethical principles of respect for human autonomy, prevention of harm, fairness and explicability are unlikely to be affected by an underwater marine mapping system.</p> <p>d) Is not correct. Such an app has the potential to be unfair to vulnerable groups, such as those with disabilities and it may also create unwanted pressure on employees.</p>	AI-2.5.1	K2	1
9	c	<p>a) Is not correct. Understanding the underlying technology is considered to be interpretability rather than explainability.</p> <p>b) Is not correct. Understanding how the loan system made a decision is more likely to be explainability.</p> <p>c) Is correct. Understanding the technology underlying the oncology system is likely to be interpretability.</p> <p>d) Is not correct. The system responding correctly is simply referring to functional correctness – not transparency, interpretability or explainability.</p>	AI-2.7.1	K2	1
10	d	<p>a) Is not correct. With unsupervised machine learning there are no labels.</p> <p>b) Is not correct. A reinforcement learning system uses a fitness function to drive its learning.</p> <p>c) Is not correct. A self-learning reinforcement system uses rewards to teach itself.</p> <p>d) Is correct. With unsupervised machine learning there are no labels and patterns are determined from the data itself.</p>	AI-3.1.2	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
11	b	<p>Considering the given descriptions:</p> <p>I. Model performance is checked using validation data. This is carried out during model evaluation.</p> <p>II. The origin of the test data used to test the model is identified. This is carried out as part of data preparation.</p> <p>III. The tuned model is made ready for its target hardware. This is carried out as part of deployment.</p> <p>IV. Test data are used to ensure the agreed performance criteria are met. This is carried out as part of testing the model.</p> <p>V. The model is created from source code. This is carried out as part of the build and compile model activity.</p> <p>VI. The critical data features are identified. This is carried out as part of data preparation.</p> <p>a) Is not correct. b) Is correct. c) Is not correct. d) Is not correct.</p>	AI-3.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
12	a	<p>a) Is correct. It is unlikely that the ML algorithm will be trained on the mobile device.</p> <p>b) Is not correct. If the health monitoring system is embedded and needs to be retrained, we would likely set a maximum duration for this as the patient must either attend a facility for it to happen or be unmonitored while it is being done remotely.</p> <p>c) Is not correct. The number of features expected to be used by the model will affect the choice of model.</p> <p>d) Is not correct. Knowing the number of classes for clustering is a useful input when choosing the ML model.</p>	AI-3.4.1	K2	1
13	d	<p>a) Is not correct. The test dataset typically comes from the same source as the validation dataset.</p> <p>b) Is not correct. The test dataset and the validation dataset should both have the same format.</p> <p>c) Is not correct. The test dataset is used for neither training nor for validation.</p> <p>d) Is correct. Incorporation of the test dataset in training will lead to bias in the evaluation of model.</p>	AI-4.2.1	K2	1
14	c	<p>a) Is not correct. Having insufficient data doesn't determine the likelihood of data mislabeling.</p> <p>b) Is not correct. Synthetic data does not lead to poor labeling.</p> <p>c) Is correct. Translation errors may lead to correctly labeled data in one language being mislabeled in a second language.</p> <p>d) Is not correct. Data labeling is not related to the choice of ML algorithm.</p>	AI-4.5.2	K1	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
15	a	<p>a) Is correct. The model quality depends on labeling quality. Wrong labeling leads to wrong ground truth. For incorrectly labeled data the functional performance measurements may indicate a good quality model but it would be producing wrong outputs.</p> <p>b) Is not correct. Values of the ML functional performance metrics are not dependent on the tool used to measure them.</p> <p>c) Is not correct. Validation data may or may not have bias in it.</p> <p>d) Is not correct. Data transformation is often performed and it does not necessarily impact the quality of the model. Wrong transformations may result in data quality issues and subsequent model quality issues, but a general statement relating data transformation to poor model quality cannot be made.</p>	AI-5.3.1	K2	1
16	a	<p>a) Is correct. Benchmark suites indicate the training time.</p> <p>b) Is not correct. Benchmark suites do not indicate the test time.</p> <p>c) Is not correct. Benchmark suites do not indicate the validation time.</p> <p>d) Is not correct. Benchmark suites do not indicate the deployment time.</p>	AI-5.5.1	K2	1
17	b	<p>a) Is not correct. Component testing is applied to non-model components and is conventional testing.</p> <p>b) Is correct. Input data testing is performed for testing bias, among other things.</p> <p>c) Is not correct. While independent bias testing can be also done as part of system testing, ideally it should be done before training the model.</p> <p>d) Is not correct. Model testing is used to check that the model alone meets any specified requirements, e.g., ML functional performance criteria and non-functional criteria.</p>	AI-7.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
18	c	<ul style="list-style-type: none"> a) Is not correct because non-functional requirements are a part of the documentation. b) Is not correct because the documentation of interfaces supports the white-box testing of component interactions. c) Is correct because bias testing on the data on the basis of its source and meta-data is possible. d) Is not correct because self-adapting AI systems rarely provide the documentation of the changes they make to themselves. 	AI-7.5.1	K2	1
19	c	<ul style="list-style-type: none"> a) Is not correct as AI effect refers to change in the perception of what is AI over time and not the accuracy of a given solution. b) Is not correct. An adversarial attack is where an attacker subtly perturbs valid inputs that are passed to the trained model to cause it to provide incorrect predictions. In this case the data is being gathered from the purchase history and not direct inputs from the users. Hence the chances of this being an adversarial attack are low. c) Is correct. This is an example of a model providing reduced accuracy because of changes in customer behavior. d) Is not correct. Fairness is using positively biased data for training which is not true in this case as we are dealing with a live system which was performing well in the past. 	AI-7.6.1	K2	1
20	a	<ul style="list-style-type: none"> a) Is correct. This is relevant to testing a system's autonomy because it may be required to test over a specific period of time to check how often the system requires intervention. b) Is not correct. A system can be autonomous and very inaccurate. c) Is not correct. A system can be autonomous and not adaptable. d) Is not correct. Static analysis of training data is unlikely to be related to testing for autonomy. 	AI-8.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
21	c	<ul style="list-style-type: none"> a) Is not correct. It is normally possible to specify expected results, but this may require the inclusion of a tolerance or making multiple test runs. b) Is not correct. Non-deterministic systems do not necessarily have explainability problems. c) Is correct. This is a valid description of a non-deterministic system, and, as a result tests may need to be run several times to give a statistically valid test result. d) Is not correct. There is no correlation between bias and non-deterministic systems. 	AI-8.2.1	K2	1
22	b	<ul style="list-style-type: none"> a) Is not correct. Methods for testing explainability of deep neural networks, which are complex to understand, require execution of the system. b) Is correct. Testing transparency does not require execution of the system under test, as it only requires comparing the documentation to the implementation. c) Is not correct. Automation bias requires execution to provide results, and determine how they are processed by users. d) Is not correct. Accuracy cannot be verified without executing the system under test. 	AI-8.6.1	K2	1
23	b	<ul style="list-style-type: none"> a) Is not correct. Data does not cause a test oracle problem because it does not mean it is difficult to determine the expected result. b) Is correct. If the ground truth of a predicted system is unknown, that is likely to cause a test oracle problem. c) Is not correct. Autonomy is unrelated to test oracles. d) Is not correct. Knowing how the system was implemented is not generally required to determine the expected result. 	AI-8.7.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
24	c	<ul style="list-style-type: none"> a) Is not correct. Back-to-back testing, and not A/B testing, is known as differential testing. b) Is not correct. A/B testing can be used for both simple and complex ML models. c) Is correct. A/B testing requires well-defined acceptance criteria to choose between the two models being tested. d) Is not correct. A/B testing doesn't help in writing test cases. 	AI-9.4.1	K2	1
25	c	<ul style="list-style-type: none"> a) Is not correct. Running rare scenarios in a virtual test environment is useful, especially when the cost of running the tests in the real world would be prohibitive. b) Is not correct. Running tests for a money market trading system on a virtual test environment is sensible as we could not run these tests on the real market due to the potential for losing money. c) Is correct. Running a fruit picker faster in a virtual environment will not help to determine if fruit is undamaged in real-time production because checking damage to fruit in the virtual environment would be far more difficult than checking on damage to real fruit in a real (non-virtual) environment. d) Is not correct. It is good practice to run dangerous test scenarios for an autonomous car in a virtual test environment on safety grounds. 	AI-10.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
26	a	<p>a) Is correct. Image recognition in the form of classification models is used to recognize objects in GUIs.</p> <p>b) Is not correct. Probabilistic software engineering can be used for estimating the probability of certain things occurring,</p> <p>c) Is not correct. Search-based software engineering techniques are used for reducing a large problem space. These may be useful for guessing where the image objects are located, but not recognizing them as images.</p> <p>d) Is not correct. Clustering is a type of algorithm, not a type of AI software engineering technology.</p>	AI-11.1.1	K2	1
27	c	<p>a) Is not correct. Visual testing makes use of images and does not use object IDs.</p> <p>b) Is not correct. Visual testing makes use of images but does not perform pixel by pixel comparison. It uses AI to do the comparison.</p> <p>c) Is correct. Visual Testing can find overlapping user interface elements similar to a human tester.</p> <p>d) Is not correct. Visual testing works even when the layout changes.</p>	AI-11.6.1	K2	1