Sample Exam – Answers
Sample Exam set A
Version 1.3

ISTQB® Certified Tester AI Testing Syllabus

Compatible with Syllabus version 1.0

International Software Testing Qualifications Board

Copyright Notice

Document Responsibility

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

Acknowledgements

Revision History

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

Version	Date	Remarks
1.3	February 9, 2026	Align version number with Questions file
1.2	December 8, 2024	Correction to Answer: #25
1.1	December 4, 2023	Correction of Answer: #19
1.0	October 1, 2021	Release for GA

<div style="border: 1px solid black;">

Table of Contents

</div>

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	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' 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 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 d) Is not correct. The gullibility of cinema goers believing killer robots will take over the world is not the 'AI Effect'	AI-1.1.1	K2	1
2	c	a) Is not correct. Support vector machines are a form of machine learning b) Is not correct. Decision trees are a form of machine learning 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 d) Is not correct. Bayesian optimization is a form of machine learning	AI-1.4.1	K1	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
3	b	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 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 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 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	AI-1.6.1	K2	1
4	b	a) Is not correct. The question mentions the good quality of the pre-trained model, so this risk should be negligible b) Is correct. The data used to train the model should be similar to the data used for making the predictions c) Is not correct. Performance does not appear to be an issue in this situation 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	AI-1.8.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
5	a	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	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
7	b	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' 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 c) Is not correct. Reward hacking is not a form of paying AI developers d) Is is not correct. Some game-playing AI-based systems are driven by a reward function, but this is not known as 'reward hacking'	AI-2.6.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points

8	a	Considering the given attributes:
		  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
		Therefore, I and IV are the attributes that are MOST problematic for safety-related systems.
		a) Is correct
		b) Is not correct
		c) Is not correct
		d) Is not correct
			AI-2.8.1	K1	1

9	b	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

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
10	a	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
11	c	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 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 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 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	AI-3.3.1	K3	2

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
12	c	a) Is not correct. The model performs well on validation data, so it is not a case of underfitting b) Is not correct. Concept drift refers to changes after the model training and validation stage c) Is correct. The bad performance on test data and good on validation data suggests overfitting 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	AI-3.5.1	K2	1
13	c	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

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
14	d	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
15	b	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. Mislabeled 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	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

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
17	c	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	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
19	b	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 b) Is correct. Precision should be used because the cost of false-positives (overfeeding the dog) is high (serious health issues) 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 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	AI-5.4.1	K4	2

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
20	b	a) Is not correct. A neural network does not have a hierarchical structure 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 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) 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)	AI-6.1.1	K2	1
21	b	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 b) Is correct. Threshold coverage measures the proportion of neurons activated during testing with a value greater than a preset threshold value 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) 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	AI-6.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
22	d	a) Is not correct. This is a specific requirement with a test oracle, so should not usually cause a testing challenge b) Is not correct. This may be a difficult requirement to achieve but should not create a testing challenge c) Is not correct. This is a testable requirement d) Is correct. This requirement is extremely complex to test without defining all human emotions and how the system might mimic them	AI-7.1.1	K2	1
23	a	a) Is correct. Sourcing data for AI systems that use large quantities of high-velocity data can be difficult b) Is not correct. Sourcing consistent data from multiple sources can be difficult c) Is not correct. Sourcing data separately is good practice as it prevents common failures with the data scientists d) Is not correct. Sourcing data from public websites is straightforward	AI-7.3.1	K1	1
24	d	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

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
25	b, e	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 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
26	c	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

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
27	a	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
28	b	a) Is not correct. Bias does not usually relate to system complexity 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 c) Is not correct. Ethics is not usually related to AI-based system complexity d) Is not correct. Difficulty finding representative data to train a model is not related to testing or AI-based system complexity	AI-8.5.1	K2	1

29	b, c	Considering each of the attributes:	AI-8.8.1	K4	2
		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 Thus, bias and explainability are the two most important attributes for the given scenario a) Is not correct b) Is correct c) Is correct d) Is not correct e) Is not correct			

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
30	b	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	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	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	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). 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). 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. Therefore, T1 and T2 combined should contain all the cameras from the source test case, but with no overlap between the two. a) Is not correct b) Is correct c) Is not correct d) Is not correct	AI-9.5.1	K3	2
34	a	a) Is correct. This is Exploratory Data Analysis which is an exploratory method b) Is not correct. This is scripted testing c) Is not correct. This is checklist-based testing d) Is not correct. Calculating ML functional performance metrics is not exploratory testing	AI-9.6.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
35	b, c	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 b) Is correct. Back-to-back testing uses a similar product as a pseudo-oracle for testing c) Is correct. Adversarial testing is important here as it being used for very important purpose and adversarial data can cause harm 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 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	AI-9.7.1	K4	2
36	a	a) Is correct. Explainability mechanism may need to be provided for AI environments b) Is not correct. Simulators and virtual environments are often required for conventional systems c) Is not correct. Large amount of data may be required for conventional systems, as well d) Is not correct. GPUs may be required for many other systems as well, for example, games	AI-10.1.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
37	b	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 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 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 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	AI-11.2.1	K2	1
38	c	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 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 c) Is correct. Web server logs may reflect production use of the system and provide a way for AI to generate tests 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	AI-11.3.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
39	b	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 b) Is correct. Optimization of regression test suites is performed by analyzing information on previous test executions 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 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	AI-11.4.1	K2	1
40	b	a) Is not correct. While natural language processing is an AI application, it is not used for defect prediction 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 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 d) Is not correct. Defect prediction does not involve scanning of code using rules. This is static analysis	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	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 b) Is not correct. Singularity is that point when AI has surpassed humans – not when they are at similar levels 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 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	AI-1.2.1	K2	1
2	b	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 b) Is correct. There is no evidence that this system implements AI c) Is not correct. This system appears to have programmed itself by listening to recordings; therefore, it is likely to be AI-based 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	AI-1.3.1	K2	1
3	d	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

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
4	c	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 AIaaS 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 AIaaS c) Is correct. The exclusive algorithm cannot be AIaaS, however the pricing part is generic and could well be provided as AIaaS 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 AIaaS	AI-1.7.1	K2	1
5	b	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
6	d	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 b) Is not correct. Supervised learning is dependent on labelled data, but flexibility is not a characteristic associated with reading poorly labelled data 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 d) Is correct. Flexibility is a characteristic associated with the ability of systems to be used in contexts outside the original requirements	AI-2.1.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
7	d	a) Is not correct. Self-learning systems that work in the same operational environment could still optimize themselves and so change their behavior b) Is not correct. AI-based systems that change themselves to adapt to changes in their environment could still cause negative side-effects 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 d) Is correct. If the self-learning system physically interacts with people, then any changes it makes to itself could potentially harm people	AI-2.3.1	K2	1
8	c	a) Is not correct. This game is aimed at children, a vulnerable group, and may give them an unfavorable view of democracy 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 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 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	AI-2.5.1	K2	1
9	c	a) Is not correct. Understanding the underlying technology is considered to be interpretability rather than explainability b) Is not correct. Understanding how the loan system made a decision is more likely to be explainability c) Is correct. Understanding the technology underlying the oncology system is likely to be interpretability d) Is not correct. The system responding correctly is simply referring to functional correctness – not transparency, interpretability or explainability	AI-2.7.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
10	d	a) Is not correct. With unsupervised machine learning there are no labels b) Is not correct. A reinforcement learning system uses a fitness function to drive its learning c) Is not correct. A self-learning reinforcement system uses rewards to teach itself d) Is correct. With unsupervised machine learning there are no labels and patterns are determined from the data itself	AI-3.1.2	K2	1

11	b	Considering the given descriptions:
		  I. Model performance is checked using validation data. This is carried out during model evaluation
		  II. The origin of the test data used to test the model is identified. This is carried out as part of data preparation
		  III. The tuned model is made ready for its target hardware. This is carried out as part of deployment
		  IV. Test data are used to ensure the agreed performance criteria are met. This is carried out as part of testing the model
		  V. The model is created from source code. This is carried out as part of the build and compile model activity
		  VI. The critical data features are identified. This is carried out as part of data preparation
		Thus:
		a) Is not correct
		b) Is correct
		c) Is not correct
		d) Is not correct
			AI-3.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
12	a	a) Is correct. It is unlikely that the ML algorithm will be trained on the mobile device 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 c) Is not correct. The number of features expected to be used by the model will affect the choice of model d) Is not correct. Knowing the number of classes for clustering is a useful input when choosing the ML model	AI-3.4.1	K2	1
13	d	a) Is not correct. The test dataset typically comes from the same source as the validation dataset b) Is not correct. The test dataset and the validation dataset should both have the same format c) Is not correct. The test dataset is used for neither training nor for validation d) Is correct. Incorporation of the test dataset in training will lead to bias in the evaluation of model	AI-4.2.1	K2	1
14	c	a) Is not correct. Having insufficient data doesn't determine the likelihood of data mislabeling b) Is not correct. Synthetic data does not lead to poor labeling c) Is correct. Translation errors may lead to correctly labeled data in one language being mislabeled in a second language d) Is not correct. Data labeling is not related to the choice of ML algorithm	AI-4.5.2	K1	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
15	a	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 b) Is not correct. Values of the ML functional performance metrics are not dependent on the tool used to measure them c) Is not correct. Validation data may or may not have bias in it 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	AI-5.3.1	K2	1
16	a	a) Is correct. Benchmark suites indicate the training time b) Is not correct. Benchmark suites do not indicate the test time c) Is not correct. Benchmark suites do not indicate the validation time d) Is not correct. Benchmark suites do not indicate the deployment time	AI-5.5.1	K2	1
17	b	a) Is not correct. Component testing is applied to non-model components and is conventional testing b) Is correct. Input data testing is performed for testing bias, among other things 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 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	AI-7.2.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
18	c	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	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	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	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	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	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	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	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
26	a	a) Is correct. Image recognition in the form of classification models is used to recognize objects in GUIs b) Is not correct. Probabilistic software engineering can be used for estimating the probability of certain things occurring 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 d) Is not correct. Clustering is a type of algorithm, not a type of AI software engineering technology	AI-11.1.1	K2	1

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points
27	c	a) Is not correct. Visual testing makes use of images and does not use object IDs 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 c) Is correct. Visual Testing can find overlapping user interface elements similar to a human tester d) Is not correct. Visual testing works even when the layout changes	AI-11.6.1	K2	1