A4Q

AI and Software Testing Foundation

Mock Exam V1.0

Released

Version 2019

Alliance for Qualification

Revision History

Date	Version	Comment
2019/06/27	0.3.1	Updates to Chapter 3 questions
2019/06/28	0.3.2	Updated to Chapter 2 justifications
2019/06/29	0.3.3	Updates to Chapter 1 questions & justifications
2019/06/29	0.3.4	Updates to Chapter 3 justifications
2019/09/30	0.3.6	Updated to Chapter 2 following review
2019/10/04	0.4.0	Updates to Chapter 1, 2 and 3 for bug fixing
2019/10/05	0.4.2	Updates to Chapter 1, 2 and 3 for bug fixing
2019/10/06	1.0	Final for release.

1.   Questions

Question 1   Keywords   K1

What are clusters?

A.   A type of testing that is run with two equivalent samples, changing one variable.
B.   Individuals that belong to the same group or share some characteristic trait.
C.   A connection or link between things.
D.   Components of a system which enable it to influence the environment.

Question 2   AI-1.2.1   K1

Among the following proposals, from which period in the history of AI does the perceptron date?

A.   After 2010
B.   Between 1960 and 1979
C.   Between 1980 and 2009
D.   Between 1940 and 1959

Question 3   AI-1.6.3   K1

Which one of the following is NOT an example of ethical issues in AI?

A.   Fairness and discrimination
B.   Very long computation time for learning
C.   Transparency
D.   Reinforcement of existing bias

Question 4   AI-1.1.2   K2

Which one of these sentences BEST describes what the Turing test is?

A.   The Turing test aims to measure the accuracy of a Machine Learning model
B.   The Turing test is based on an oral conversation between an AI system and a human
C.   The Turing test consists of putting a human being in blind verbal conversation with a computer and another human
D.   The Turing test measures the IQ of an AI system

Question 5   AI-1.4.1   K2

Given the following statements about the way a machine learning system is trained:

1. Supervised learning uses reward and punishment mechanisms to organize the learning process.
2. In reinforcement learning, the machine learning model interacts with the environment.
3. Unsupervised learning use labelled data during training.
4. Semi-supervised learning uses unlabeled data during training and labeled data during validation.

Which of the following CORRECTLY shows which are true and false?

A.   A. True – 1;   False – 2, 3, 4
B.   B. True – 2;   False – 1, 3, 4
C.   C. True – 1, 2, 4;   False – 3
D.   D. True – 1, 4;   False – 2, 3

Question 6   AI-1.6.1   K2

In the context of a banking project to estimate loan risk, machine learning is used for risk prediction. The data scientist on the project starts by choosing an ML algorithm for supervised learning, initiates its implementation and evaluation on the available data and then to deploy the ML component.

Which of the following is the MOST likely missing ML activity in this case?

A.   Integrating the machine learning model with other software
B.   Data understanding, preparation and cleaning
C.   Configuring the hyperparameters of the chosen ML model
D.   Unit testing of the whole system

Question 7   AI-1.3.1   K2

Which one of these sentences related to mathematical logic and inference is true?

A.   Propositional logic is based on variables and quantifiers
B.   Predicate logic supports a stronger expression power than propositional logic
C.   Fuzzy logic uses two values of truths ("true" and "false")
D.   Predicate logic provides a finite deterministic decision procedure

Question 8   AI-1.4.3   K2

Given the following machine learning algorithms and types of learning:

a.   q-Learning
b.   Autoencoder
c.   K-means clustering
d.   Support-vector machines

1.   Supervised learning
2.   Unsupervised learning
3.   Reinforcement learning

Which of the following BEST matches machine learning algorithms and types of learning?

A.   a-2, b-3, c-1, d-1
B.   a-2, b-1, c-2, d-2
C.   a-3, b-2, c-1, d-1
D.   a-3, b-2, c-2, d-1

Question 9   AI-1.5.1   K2

Which one of these sentences is FALSE regarding how Bayesian belief networks work?

A.   The inference model of Bayesian belief networks is based on estimating the probability of "A given B" as a function of the probability of "B given A" and the probability of A
B.   Bayesian networks can be trained using unsupervised learning, then inferring probabilities on new data
C.   A Bayesian network is a directed acyclic graph
D.   After training, Bayesian networks may be used to classify data in anomaly detection

Question 10   AI-1.5.3   K2

Which one of the following statements is the BEST description of the SVM algorithm?

A.   The SVM algorithm computes an hyperplane to separate data according to a distance function
B.   The SVM algorithm uses a probabilistic Bayesian classification to perform supervised learning
C.   The SVM algorithm learns a weight vector during training session
D.   The SVM algorithm requires that the number of clusters be initialized initially

Question 11   AI-1.5.2   K3

Body mass index (BMI) category statistical data on a sample of a population practicing a sport, and ranging from 20 to 80 years old spread over 3 age groups (20-35, 36-60 and 61-80), made it possible to compute the following probabilities:

- P(BMI category is between 18.5 – 24.9) = 50%
- P(BMI category is between 25 – 29.9) = 30%
- P(BMI category is between 30 – 34.9) = 20%

- P(Age group = 36-60 | BMI category is between 18.5 – 24.9) = 30%
- P(Age group = 36-60 | BMI category is between 25 – 29.9) = 40%
- P(Age group = 36-60 | BMI category is between 30 – 34.9) = 30%

- P(Type of sport = Running | BMI category is between 18.5 – 24.9) = 60%
- P(Type of sport = Running | BMI category is between 25 – 29.9) = 30%
- P(Type of sport = Running | BMI category is between 30 – 34.9) = 10%

We want to classify Mary, who is 43 years old and runs twice a week, with respect to her MOST probable BMI category. Using a Naïve Bayes algorithm and previous calculations, which of the following classifications is correct?

A.   Mary has no likelihood of being in any of these BMI categories listed
B.   Mary's BMI category is most likely 25 – 29.9
C.   Mary's BMI category is most likely 30 – 34.9
D.   Mary's BMI category is most likely 18.5-24.9

Question 12   AI-1.5.5   K3

After training, the calculation of weights on a perceptron model with two inputs x1 and x2 is as follows: w1= -10 and w2=10, with a constant w0=10.

The activation function for the output value y is defined as: y=1 if w0 + x1*w1 + x2*w2 >= 0 and otherwise y=0.

From this perceptron model, we want to classify two data vectors: A=(2,3) and B=(1,0). Which of the following solutions is correct?

A.   y=1 for vector A, and y=1 for vector B
B.   y=1 for vector A, and y=0 for vector B
C.   y=0 for vector A, and y=1 for vector B
D.   y=0 for vector A, and y=0 for vector B

Question 13   Keywords   K1

Which of these are actuators?

A.   A measurement device
B.   Something that has an effect in the real-world
C.   Real world sensors
D.   The degree to which the system knows how its environment will be behave and respond.

Question 14   AI-2.1.1   K1

What is a non-deterministic system?

A.   A system that cannot determine whether the result proposed by an expert is correct
B.   A system that cannot be tested
C.   A system that given the same inputs and initial state, may produce different output
D.   A system that always gives the same result

Question 15   AI-2.1.4   K1

Which two of the below aspects of AI systems can make it difficult to design tests?

A.   They are often trained on data or knowledge rather than specified in detail
B.   They can be probabilistic and non-deterministic
C.   There may be multiple agents in the environment
D.   The mathematics of machine learning model is difficult to understand
E.   The bias/variance trade-off is always present

Question 16   AI-2.1.5   K1

A weak test basis or test oracle can cause which of the following problems during test planning and execution?

A.   Challenges in raising defect reports for the correct components
B.   Difficulties understanding the required test levels
C.   Complicated architecture which can't be easily tested
D.   Difficulties in determining the expected results of tests

Question 17   AI-2.1.2   K1

What is a probabilistic system?

A.   A system where the expected results cannot be accurately described without using probabilities
B.   A system that uses machine learning
C.   A system that uses probabilities to determine how to respond to an input
D.   A system that uses algorithms to simulate intelligence

Question 18 AI-2.1.7   K2

Which of the below are NOT reasons that machine learning models can have defects as a result of the data chosen to train them?

A. The trade-off between consistency and accuracy
B. Drift
C. Incorrect labelling of data
D. Self-optimization

Question 19   AI-2.2.1   K2

Which of the below is NOT a way to train a machine learning model which is designed to predict what customers will purchase?

A. By labeling customer records based on whether they have specified interest in buying a product.
B. Through users interacting with the system through a web browser.
C. Specifying an explicit rule that the model should predict customers from France will purchase wine.
D. Analyzing prior customer purchases.

Question 20 AI-2.4.2   K2

Which of these are good examples of acceptance criteria for AI systems?

A. More than 80% of data has been held back as testing data
B. A classification algorithm for prediction is desired to achieve no more than 12% false positive errors and no false negative errors, but up to 15%/5% will be acceptable.
C. The test sample size is 20% of the total population size.
D. The system should not incorrectly predict that a record belongs to a target class in more than 5% of cases.

Question 21 AI-2.4.3   K2

When testing a system where the correct answer can only be determined with very specialist knowledge, which functional testing approaches could NOT be used?

A. Metamorphic testing
B. Expert panels
C. A/B testing
D. Fuzz testing

Question 22 AI-2.4.8   K2

Which of the below is NOT a reason to conduct acceptance testing with AI systems?

A. Real users vary in how they use natural language
B. Comparing an automated solution to human users performing the same process
C. Monitoring ongoing conversion rates in production
           D. To measure automation bias

Question 23   AI-2.1.3   K2

Which two of these examples are non-testable systems?

A. A system that produces more output than can be practically consumed
B. A system that does not give consistent results
C. Machine learning
D. A system that requires 100 real users to test simultaneously
E. A system that predicts an event in the future, that humans cannot reasonably predict

Question 24 AI-2.1.9   K2

Which of the below is a reason why drift would occur?

A. A change in the relationship between inputs and outputs of an algorithm
B. The absence of a test oracle
C. Incorrect training data used
D. Test reporting

Question 25 AI-2.3.2   K2

What should be considered when specifying an environment used to test an intelligent agent?

A. Test oracles
B. Knowledge
C. Overfitting
D. Supervised learning

Question 26 AI-2.4.4   K3

After completing test design, you are about to start testing a new robotic AI system for the first time. Why would metamorphic testing NOT be appropriate in this context?

A. It cannot be used in the first test execution
B. Metamorphic testing because it changes the AI system through the act of testing
C. It is mostly used for search algorithms
D. It relies on the relationship between the test environment and the system

Question 27 AI-2.4.6   K3

Which of the below test techniques would be most appropriate for acceptance testing a purchase recommender system?

A. A/B testing
B. Model training
C. Metamorphic testing
D. Component testing

Question 28   Keywords   K1

What is a test oracle?

A. A source of data used for testing.
B. An AI-related testing approach.
C. A means to determine the passing or failure of a test case.
D. An open source testing tool from Oracle, Inc.

Question 29   AI-3.4.1   K1

Which of the following statements is correct?

A. Monkey testing can NOT be combined with AI approaches.
B. The implicit test oracle can NOT be used for monkey testing.
C. In a random string, the probability for any given sequence of characters to occur is non-zero.
D. Monkey testing will always produce the same output.

Question 30   AI-3.3.1   K1

What is a typical problem when generating tests on component level (unit tests)?

A. The implicit test oracle is often defunct.
B. There is a clear system boundary.
C. The output format is much less formal.
D. The use of external data to validate the quality of a system.

Question 31   AI-3.1.1   K2

What are possible test oracles?

i.   Specified test oracles
ii.   Well-defined test oracles
iii.   Derived test oracles
iv.   Manual test oracles
v.   Implicit test oracles

A. ii, iii true   i, iv, v false
B. i, iii, v true   ii, iv false
C. i, iv true   ii, iii, v false
D. ii, iv, v true   i, iii false

Question 32   AI-3.2.3   K2

Which of the following sentences is NOT correct?

A.   Classification and labeling algorithms are suited for a wide range of test data generation tasks.
B.   A genetic algorithm can be used to generate structured input data.
C.   Fuzz testing often is an appropriate method to generate test data.
D.   A special form of neural networks can be used to generate images.

Question 33   AI-3.4.3   K2

Match each item of the two lists to create correct statements?

1.   An unsupervised learning method
2.   Code coverage
3.   A supervised learning method
4.   Monkey testing

a.   can be used for test selection.
b.   requires the availability of past test performance.
c.   cannot be used for test selection.
d.   can be used as additional input.

A.   1a, 2b, 3d and 4c
B.   1a, 2c, 3d and 4b
C.   1b, 2c, 3a and 4d
D.   1a, 2d, 3b and 4c

Question 34   AI-3.3.2   K2

Why is test generation on the system level easier than on the component level?

1. The correct initial state
2. The input
3. An implicit test oracle
4. The output format

a. can be used.
b. can be much less formal.
c. is created by the system itself.
d. consists of primitive types.

A. 1a, 2b, 3c and 4d
B. 1c, 2d, 3a and 4b
C. 1b, 2c, 3a and 4d
D. 1a, 2d, 3b and 4c

Question 35   AI-3.4.4   K2

What information can be used as input for an AI approach to identifier selection?

A. Code coverage
B. Screenshots of elements
C. Confusion matrix
D. Confidence in identifiers

Question 36   AI-3.4.6   K2

What are the goals of visual test automation?

A. Identifying UI/UX problems
B. Identifying functional regressions
C. Identifying accessibility problems
D. Identifying visual regressions

Question 37   AI-3.5.1   K2

Which TWO of the following effects are emphasized by AI-based testing approaches?

A.   Defect clustering
B.   Pesticide paradox
C.   Absence-of-errors fallacy
D.   Self-fulfilling prophecy
E.   No Free Lunch paradox

Question 38   AI-3.4.2   K2

Which approaches can be used for test generation?

 i.   Genetic algorithms
 ii.   K-means clustering
 iii.   Neural networks
 iv.   Principal component analysis
 v.   Support vector machines

A.   ii, iii true   i, iv, v false
B.   i, iii true   i, iv, v false
C.   i, iv true   ii, iii, v false
D.   ii, iv, v true i, iii false

Question 39   AI-3.4.5   K2

Which TWO of the following AI approaches can be used for visual test automation?

A.   Image recognition
B.   Genetic algorithms
C.   Cross-platform testing
D.   Supervised learning algorithms
E.   Unsupervised learning algorithms

Question 40   AI-3.5.2   K3

Which of the following combination represents relevant aspects and questions to ask during a test generation tool evaluation?

1. Test case relevance
2. Defect relevance
3. Test case explosion
4. Maintainability

a. Which counter measures are in place?
b. How high is the expected customer impact?
c. Is there monitoring of effort and added benefit?
d. Are business use cases and associated risks covered?

A. 1d, 2b, 3a and 4c
B. 1a, 2c, 3d and 4b
C. 1b, 2c, 3a and 4d
D. 1c, 2d, 3b and 4a

2.   Answer Key

Question Number	Correct Answer	LO	K-Level
1	B	Keywords	1
2	D	AI-1.2.1	1
3	B	AI-1.6.3	1
4	C	AI-1.1.2	2
5	B	AI-1.4.1	2
6	B	AI-1.6.1	2
7	B	AI-1.3.1	2
8	D	AI-1.4.3	2
9	B	AI-1.5.1	2
10	A	AI-1.5.3	2
11	D	AI-1.5.2	3
12	A	AI-1.5.5	3
13	B	Keywords	1
14	C	AI-2.1.1	1
15	A, B	AI-2.1.4	1
16	D	AI-2.1.5	1
17	A	AI-2.1.2	1
18	D	AI-2.1.7	2
19	C	AI-2.2.1	2
20	B	AI-2.4.2	2
21	D	AI-2.4.3	2
22	C	AI-2.4.8	2
23	A, E	AI-2.1.3	2
24	A	AI-2.1.9	2
25	B	AI-2.3.2	2
26	A	AI-2.4.4	3
27	A	AI-2.4.6	3
28	C	Keywords	1
29	C	AI-3.4.1	1
30	A	AI-3.3.1	1
31	B	AI-3.1.1	2
32	A	AI-3.2.3	2
33	D	AI-3.4.3	2
34	B	AI-3.3.2	2
35	B	AI-3.4.4	2
36	D	AI-3.4.6	2
37	B, D	AI-3.5.1	2
38	B	AI-3.4.2	2
39	A, D	AI-3.4.5	2
40	A	AI-3.5.2	3

3.   Questions with Justifications

Question 1   Keywords   K1

What are clusters?

A.   A type of testing that is run with two equivalent samples, changing one variable.
B.   Individuals that belong to the same group or share some characteristic trait.
C.   A connection or link between things.
D.   Components of a system which enable it to influence the environment.

Justification

A.   Incorrect. This is the definition of an A/B test.
B.   Correct. This is part of the Glossary definition of a cluster.
C.   Incorrect. This is part of the definition of correlation.
D.   Incorrect. This is the definition actuators.

Question 2   AI-1.2.1   K1

Among the following proposals, from which period in the history of AI does the perceptron date?

A.   After 2010
B.   Between 1960 and 1979
C.   Between 1980 and 2009
D.   Between 1940 and 1959

Justification

A.   Incorrect. The perceptron was proposed in 1958.
B.   Incorrect. The perceptron was proposed in 1958.
C.   Incorrect. The perceptron was proposed in 1958.
D.   Correct. The perceptron was proposed in 1958.

Question 3   AI-1.6.3   K1

Which one of the following is NOT an example of ethical issues in AI?

A.   Fairness and discrimination
B.   Very long computation time for learning
C.   Transparency
D.   Reinforcement of existing bias

Justification

A. Incorrect. Fairness and discrimination are ethical issues in AI. See Section 1.6 of the syllabus.
B. Correct. This is a technical problem. See Section 1.6 of the syllabus.
C. Incorrect. Transparency is an ethical issue in AI. See Section 1.6 of the syllabus.
D. Incorrect. Reinforcement of existing bias is an ethical issue in AI. See Section 1.6 of the syllabus.

Question 4   AI-1.1.2   K2

Which one of these sentences BEST describes what the Turing test is?

A. The Turing test aims to measure the accuracy of a Machine Learning model
B. The Turing test is based on an oral conversation between an AI system and a human
C. The Turing test consists of putting a human being in blind verbal conversation with a computer and another human
D. The Turing test measures the IQ of an AI system

 Justification

A. Incorrect. Measuring accuracy is not the goal of the Turing test.
B. Incorrect. The Turing test is limited to text message (not an oral conversation).
C. Correct. See syllabus – section 1.4.
D. Incorrect. Measuring the IQ is not the goal of the Turing test.

Question 5   AI-1.4.1   K2

Given the following statements about the way a machine learning system is trained:

1. Supervised learning uses reward and punishment mechanisms to organize the learning process.
2. In reinforcement learning, the machine learning model interacts with the environment.
3. Unsupervised learning use labelled data during training.
4. Semi-supervised learning uses unlabeled data during training and labeled data during validation.

 Which of the following CORRECTLY shows which are true and false?

A. True – 1;   False – 2, 3, 4
B. True – 2;   False – 1, 3, 4
C. True – 1, 2, 4;   False – 3
D. True – 1, 4;   False – 2, 3

Justification

Considering each statement:
1. Supervised learning uses reward and punishment mechanisms to organize the learning.
        FALSE – This is true for reinforcement learning as described in the syllabus (1.4.1).
 2. In reinforcement learning, the machine learning model interacts with the environment.
        TRUE – In reinforcement learning, the machine learning model interacts with the environment, as per syllabus (1.4.1).

3. Unsupervised learning use labelled data during training.
>   FALSE – This is true for supervised learning as described in the syllabus (1.4.1).
4. Semi-supervised learning uses unlabeled data during training and labeled data during validation.
>   FALSE - Semi-supervised Learning combines both supervised and unsupervised learning by using a set of labelled and unlabeled data for training (see syllabus 1.4.1).

Thus, option B is CORRECT.

Question 6   AI-1.6.1   K2

In the context of a banking project to estimate loan risk, machine learning is used for risk prediction. The data scientist on the project starts by choosing an ML algorithm for supervised learning, initiates its implementation and evaluation on the available data and then to deploy the ML component.

Which of the following is the MOST likely missing ML activity in this case?

A. Integrating the machine learning model with other software
B. Data understanding, preparation and cleaning
C. Configuring the hyperparameters of the chosen ML model
D. Unit testing of the whole system

 Justification

A. Incorrect. Fairness and discrimination are ethical issues in AI. See Section 1.6 of the syllabus.
B. Correct. Data understanding, preparation and cleaning are essential activities of ML projects: poor data lead to poor results. See Section 1.6 of the syllabus.
C. Incorrect. The configuration of hyperparameters is part of the implementation of the ML model so this is done in this case. See Section 1.6 of the syllabus.
D. Incorrect. Unit testing of the whole system is not the first priority here because the focus of the question is on the ML component. See Section 1.6 of the syllabus.

Question 7   AI-1.3.1   K2

Which one of these sentences related to mathematical logic and inference is true?

A. Propositional logic is based on variables and quantifiers
B. Predicate logic supports a stronger expression power than propositional logic
C. Fuzzy logic uses two values of truths ("true" and "false")
D. Predicate logic provides a finite deterministic decision procedure

 Justification

A. Incorrect. There are not quantified variables in propositional logic (see syllabus – section 1.3.1).
B. Correct. See syllabus – section 1.3.1.
C. Incorrect. Fuzzy logic uses an infinite number of truth values (see syllabus – section 1.3.1).
D. Incorrect. Predicate logic is generally undecidable (see syllabus – section 1.3.1).

Question 8   AI-1.4.3   K2

Given the following machine learning algorithms and types of learning:

a. q-Learning
b. Autoencoder
c. K-means clustering
d. Support-vector machines

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Which of the following BEST matches machine learning algorithms and types of learning?

A. a-2, b-3, c-1, d-1
B. a-2, b-1, c-2, d-2
C. a-3, b-2, c-1, d-1
D. a-3, b-2, c-2, d-1

Justification

D is CORRECT – see syllabus – section 1.4.3

Question 9   AI-1.5.1   K2

Which one of these sentences is FALSE regarding how Bayesian belief networks work?

A. The inference model of Bayesian belief networks is based on estimating the probability of "A given B" as a function of the probability of "B given A" and the probability of A
B. Bayesian networks can be trained using unsupervised learning, then inferring probabilities on new data
C. A Bayesian network is a directed acyclic graph
D. After training, Bayesian networks may be used to classify data in anomaly detection

Justification

A. Incorrect. This sentence is true (see syllabus – section 1.5.1).
B. Correct. Bayesian networks are trained with labelled data (supervised learning).
C. Incorrect. This sentence is true (see syllabus – section 1.5.1).
D. Incorrect. This sentence is true (see syllabus – section 1.5.1).

Question 10   AI-1.5.3   K2

Which one of the following statements is the BEST description of the SVM algorithm?

A. The SVM algorithm computes an hyperplane to separate data according to a distance function
B. The SVM algorithm uses a probabilistic Bayesian classification to perform supervised learning
C. The SVM algorithm learns a weight vector during training session
D. The SVM algorithm requires that the number of clusters be initialized initially

 Justification

A. Correct. See syllabus – section 1.5.3.
B. Incorrect. This is correct for the Naive Bayes classifier not the SVM algorithm.
C. Incorrect. This is correct for the perceptron not the SVM algorithm.
D. Incorrect. This is correct for the K-Means algorithm not the SVM algorithm.

Question 11   AI-1.5.2   K3

Body mass index (BMI) category statistical data on a sample of a population practicing a sport, and ranging from 20 to 80 years old spread over 3 age groups (20-35, 36-60 and 61-80), made it possible to compute the following probabilities:

- P(BMI category is between 18.5 – 24.9) = 50%
- P(BMI category is between 25 – 29.9) = 30%
- P(BMI category is between 30 – 34.9) = 20%

- P(Age group = 36-60 | BMI category is between 18.5 – 24.9) = 30%
- P(Age group = 36-60 | BMI category is between 25 – 29.9) = 40%
- P(Age group = 36-60 | BMI category is between 30 – 34.9) = 30%

- P(Type of sport = Running | BMI category is between 18.5 – 24.9) = 60%
- P(Type of sport = Running | BMI category is between 25 – 29.9) = 30%
- P(Type of sport = Running | BMI category is between 30 – 34.9) = 10%

We want to classify Mary, who is 43 years old and runs twice a week, with respect to her MOST probable BMI category. Using a Naïve Bayes algorithm and previous calculations, which of the following classifications is correct?

A. Mary has no likelihood of being in any of these BMI categories listed
B. Mary's BMI category is most likely 25 – 29.9
C. Mary's BMI category is most likely 30 – 34.9
D. Mary's BMI category is most likely 18.5-24.9

Justification

Using the Naïve Bayes algorithm, we can compute:

- P(Mary's BMI category is 18.5-24.9) $\propto$ 50% * 30% * 60% = 9 %
- P(Mary's BMI category is 25 – 29.9) $\propto$ 30% *40% * 30% = 3.6 %
- P(Mary's BMI category is 30 – 34.9) $\propto$ 20% * 30% * 10% = 0.6 %

where $\propto$ denotes proportionality.

Thus, option D is correct.

Question 12   AI-1.5.5   K3

After training, the calculation of weights on a perceptron model with two inputs x1 and x2 is as follows: w1= -10 and w2=10, with a constant w0=10.

The activation function for the output value y is defined as: y=1 if w0 + x1*w1 + x2*w2 >= 0 and otherwise y=0.

From this perceptron model, we want to classify two data vectors: A=(2,3) and B=(1,0). Which of the following solutions is correct?

A. y=1 for vector A, and y=1 for vector B
B. y=1 for vector A, and y=0 for vector B
C. y=0 for vector A, and y=1 for vector B
D. y=0 for vector A, and y=0 for vector B

Justification

For vector A, w0 + x1*w1 + x2*w2 = 10 + -10*2 + 10*3 = 20, so y=1
For vector B, w0 + x1*w1 + x2*w2 = 10 + -10*1 + 10*0 = 0, so y=1
Thus, option A is correct.

Question 13   Keywords   K1

Which of these are actuators?

A. A measurement device
B. Something that has an effect in the real-world
C. Real world sensors
D. The degree to which the system knows how its environment will be behave and respond.

Justification

A. Incorrect. Per section 2.4, these are sensors.
B. Correct. These are actuators per section 2.4
C. Incorrect. Per section 2.4, these are sensors.
D. Incorrect. Per section 2.4, this is knowledge.

Question 14   AI-2.1.1   K1

What is a non-deterministic system?

A. A system that cannot determine whether the result proposed by an expert is correct
B. A system that cannot be tested
C. A system that given the same inputs and initial state, may produce different output
D. A system that always gives the same result

Justification

A. Incorrect. This would simply be a limited software system. See Section 2.1 of the syllabus.
B. Incorrect. This is a non-testable program per section 2.1.
C. Correct. This is the logical opposite of the definition of a deterministic system in section 2.1.
D. Incorrect. This is a deterministic system per section 2.1.

Question 15   AI-2.1.4   K1

Which two of the below aspects of AI systems can make it difficult to design tests?

A. They are often trained on data or knowledge rather than specified in detail
B. They can be probabilistic and non-deterministic
C. There may be multiple agents in the environment
D. The mathematics of machine learning model is difficult to understand
E. The bias/variance trade-off is always present

Justification:

A. Correct. The act of training on data makes it less practical to determine the test oracle, i.e. expected results of the system, in advance. See Section 2.1 of the syllabus.
B. Correct. Probabilistic and non-deterministic systems have a weaker test oracle.   See Section 2.1 of the syllabus.
C. Incorrect. The existing of multiple agents does not directly make it difficult to specify a test oracle. See Section 2.1 of the syllabus.
D. Incorrect. Testing machine learning does not require an understanding of the mathematics of a specific model. In the same way that it is not necessary to understand software code to test it. See Section 2.2 of the syllabus.
E. Incorrect. The bias/variance trade-off does not directly make it difficult to design tests.   See Section 2.1 of the syllabus.

Question 16   AI-2.1.5   K1

A weak test basis or test oracle can cause which of the following problems during test planning and execution?

A. Challenges in raising defect reports for the correct components
B. Difficulties understanding the required test levels
C. Complicated architecture which can't be easily tested
D. Difficulties in determining the expected results of tests

Justification:

A. Incorrect. Whilst it is difficult to justify defects without a strong test oracle, it does not affect the reporter's ability to identify the correct component. See Section 2.1 of the syllabus.
B. Incorrect. A weak test basis or oracle does not normally make it difficult to understand test levels. See Section 2.1 of the syllabus.
C. Incorrect. A weak test basis or oracle does not directly cause architectural complexity. See Section 2.1 of the syllabus.
D. Correct. It is difficult to complete test design without being able to specify the expected result.   See Section 2.1 of the syllabus.

Question 17   AI-2.1.2   K1

What is a probabilistic system?

A. A system where the expected results cannot be accurately described without using probabilities
B. A system that uses machine learning
C. A system that uses probabilities to determine how to respond to an input
D. A system that uses algorithms to simulate intelligence

Justification

A. Correct. This is the definition of probabilistic system as described in section 2.1.
B. Incorrect. Whilst machine learning systems are often probabilistic, not all probabilistic systems use machine learning.   See Section 2.1 of the syllabus.
C. Incorrect. Probabilistic refers to our ability to describe the expected behavior of a system, not how it works.   See Section 2.1 of the syllabus.
D. Incorrect. Probabilistic refers to our ability to describe the expected behavior of a system, not how it works.   See Section 2.1 of the syllabus.

Question 18 AI-2.1.7   K2

Which of the below are NOT reasons that machine learning models can have defects as a result of the data chosen to train them?

A. The trade-off between consistency and accuracy
B. Drift
C. Incorrect labelling of data
D. Self-optimization

Justification:

A. Incorrect. The data selected to train the models can affect the accuracy.   See Section 2.2 of the syllabus.
B. Incorrect. If the relationship between the inputs and outputs changes since the model was training, this can lead to defects, or incorrect behavior. See Section 2.2 of the syllabus.
C. Incorrect. Data that is labelled incorrectly will lead to inaccurate models. See Section 2.2 of the syllabus.
D. Correct. Self-optimization occurs after training. See Section 2.2 of the syllabus.

Question 19   AI-2.2.1   K2

Which of the below is NOT a way to train a machine learning model which is designed to predict what customers will purchase?

A. By labeling customer records based on whether they have specified interest in buying a product.
B. Through users interacting with the system through a web browser.
C. Specifying an explicit rule that the model should predict customers from France will purchase wine.
D. Analyzing prior customer purchases.

Justification:

A. Incorrect. Labeling data with predictions is a supervised machine learning technique as explained in Section 2.3.
B. Incorrect. Human interaction is a reinforcement learning technique as explained in Section 2.3.
C. Correct. Rule-based knowledge is symbolic AI rather than machine learning as explained in Section 1.3.
D. Incorrect. Historical data is usually used to label and train machine learning models as explained in Section 2.3.

Question 20 AI-2.4.2   K2

Which of these are good examples of acceptance criteria for AI systems?

A. More than 80% of data has been held back as testing data
B. A classification algorithm for prediction is desired to achieve no more than 12% false positive errors and no false negative errors, but up to 15%/5% will be acceptable.
C. The test sample size is 20% of the total population size.
D. The system should not incorrectly predict that a record belongs to a target class in more than 5% of cases.

Justification

A. Incorrect. This is a useful measurement of the coverage of a model training process as described in Section 2.2, however it is not a useful statement for system acceptance criteria.
B. Correct. This is a well-structured criterion include false positive and negative outcomes, and ranges. This is in line with Section 2.4.
C. Incorrect. This is a useful measurement of the coverage of a testing process as described in Section 2.5 (Statistical Significance).
D. Incorrect. This criterion does not allow for False Negatives. See the syllabus Section 2.4.

Question 21 AI-2.4.3   K2

When testing a system where the correct answer can only be determined with very specialist knowledge, which functional testing approaches could NOT be used?

A. Metamorphic testing
B. Expert panels
C. A/B testing
D. Fuzz testing

Justification

A. Incorrect. Metamorphic testing is a type of pseudo-oracle as explained in Section 2.4, and therefore could be used.
B. Incorrect. Expert panels can be used where specialist knowledge is required to determine the test result, as explained in Section 2.4.
C. Incorrect. A/B is useful for situations with oracle problems. See section 2.4.
D. Correct. Fuzz testing is not useful for solving a testing problem where the correct answer cannot be easily determined. See section 3.2.

Question 22 AI-2.4.8   K2

Which of the below is NOT a reason to conduct acceptance testing with AI systems?

A. Real users vary in how they use natural language
B. Comparing an automated solution to human users performing the same process
C. Monitoring ongoing conversion rates in production
D. To measure automation bias

Justification:

A. Incorrect. Variance in real user behavior is a good reason to conduct acceptance testing.   See Section 2.4.
B. Incorrect. Comparing human to system behavior is a good reason to conduct acceptance testing. See Section 2.4.
C. Correct. Monitoring ongoing production behavior is not generally acceptance testing.   See Section 2.4.
D. Incorrect.   Measuring automation bias is a good reason to conduct acceptance testing.   See Section 2.4.

Question 23   AI-2.1.3   K2

Which two of these examples are non-testable systems?

A. A system that produces more output than can be practically consumed
B. A system that does not give consistent results
C. Machine learning
D. A system that requires 100 real users to test simultaneously
E. A system that predicts an event in the future, that humans cannot reasonably predict

Justification:

A. Correct. If it is impractical to consume the output it will be difficult to determine if the system is passing tests.   As described in section 2.1 of the syllabus.
B. Incorrect. This is a non-deterministic system, which is not necessarily non-testable. See section 2.1 of the syllabus.
C. Incorrect. Machine Learning algorithms are testable if the expected results can be determined. See section 2.1 of the syllabus.
D. Incorrect. The difficulties of coordinating 100 users do not make the system non-testable. See section 2.1 of the syllabus.
E. Correct. An event in the future that cannot be reasonably predicted, cannot be testable, at least until the future time has passed! See section 2.1 of the syllabus.

Question 24 AI-2.1.9   K2

Which of the below is a reason why drift would occur?

A. A change in the relationship between inputs and outputs of an algorithm
B. The absence of a test oracle
C. Incorrect training data used
D. Test reporting

Justification:

A. Correct. Drift can be caused by changes in the relationship between inputs and outputs.   See section 2.1 of the syllabus.
B. Incorrect. The absence of a test oracle does not cause drift. See section 2.1 of the syllabus.
C. Incorrect. Incorrect training data will result in an inaccurate model, but this is not known as drift. See section 2.1 of the syllabus.
D. Incorrect. The act of testing a model which self-optimizes can cause drift, but not the act of reporting. See section 2.1 of the syllabus.

Question 25 AI-2.3.2   K2

What should be considered when specifying an environment used to test an intelligent agent?

A. Test oracles
B. Knowledge
C. Overfitting
D. Supervised learning

Justification

A. Incorrect. Test oracles are not part of the D-SOAKED model described in section 2.3.
B. Correct. Knowledge is part of the D-SOAKED model described in section 2.3.
C. Incorrect. Overfitting is not part of the environment as described in section 2.3.
D. Incorrect. Supervised learning is not part of the environment as described in section 2.3.

Question 26 AI-2.4.4   K3

After completing test design, you are about to start testing a new robotic AI system for the first time. Why would metamorphic testing NOT be appropriate in this context?

A. It cannot be used in the first test execution
B. Metamorphic testing because it changes the AI system through the act of testing
C. It is mostly used for search algorithms
D. It relies on the relationship between the test environment and the system

Justification

A. Correct.   Metamorphic testing must involve multiple executions. See section 2.4.
B. Incorrect. It is not called metamorphic testing because it changes the system under test. See Section 2.4.
C. Incorrect. Metamorphic testing can be used for many applications. See section 2.4.
D. Incorrect. Metamorphic testing does not rely on the environment. See section 2.4.

Question 27 AI-2.4.6   K3

Which of the below test techniques would be most appropriate for acceptance testing a purchase recommender system?

A. A/B testing
B. Model training
C. Metamorphic testing
D. Component testing

Justification:

A. Correct. A/B testing allows for field testing recommendation systems across different groups of users to compare their effectiveness. See section 2.4.
B. Incorrect. Model training is where the majority of the work is performed to verify the accuracy of recommendations, but acceptance testing requires the involvement of users or stakeholders, which is not usually the case in model training. See section 2.4.
C. Incorrect. Metamorphic testing is rarely relevant for acceptance testing. See section 2.4.
D. Incorrect. Component testing is rarely relevant for acceptance testing. See section 2.4.

Question 28   Keywords   K1

What is a test oracle?

A. A source of data used for testing.
B. An AI-related testing approach.
C. A means to determine the passing or failure of a test case.
D. An open source testing tool from Oracle, Inc.

Justification

A. Incorrect. Per section 3.1.1, a test oracle is related to the outcome of a test, not its input.
B. Incorrect. Per section 3.1.1, a test oracle is needed for any test and is not AI-specific.
C. Correct. Per section 3.1.1, a test oracle is a source to determine expected results to compare with the actual result of the system under test.
D. Incorrect. Similar naming seems to be purely incidental, although probably inspired by the same mythological roots.

Question 29   AI-3.4.1   K1

Which of the following statements is correct?

A. Monkey testing can NOT be combined with AI approaches.
B. The implicit test oracle can NOT be used for monkey testing.
C. In a random string, the probability for any given sequence of characters to occur is non-zero.
D. Monkey testing will always produce the same output.

Justification

A. Incorrect. As stated in 3.4.1, monkey testing CAN be combined with AI approaches.
B. Incorrect. As stated in 3.4.1, monkey testing only makes sense using the implicit test oracle.
C. Correct. See section 3.4.1.
D. Incorrect. This is the definition of a deterministic system.

Question 30   AI-3.3.1   K1

What is a typical problem when generating tests on component level (unit tests)?

A. The implicit test oracle is often defunct.
B. There is a clear system boundary.
C. The output format is much less formal.
D. The use of external data to validate the quality of a system.

Justification

A. Correct (see section 3.3.1).
B. This is not a problem with component level tests, but an advantage of system level tests (see section 3.3.2).
C. This is not a problem with component level tests, but an advantage of system level tests (see section 3.3.2).
D. This is the definition of external validity.

Question 31   AI-3.1.1   K2

What are possible test oracles?

i.   Specified test oracles
ii.   Well-defined test oracles
iii.   Derived test oracles
iv.   Manual test oracles
v.   Implicit test oracles

A.   ii, iii true   i, iv, v false
B.   i, iii, v true   ii, iv false
C.   i, iv true   ii, iii, v false
D.   ii, iv, v true   i, iii false

Justification

According to section 3.1.2, test oracles can be specified, derived or implicit. So, the answer is B. What it means for an oracle to be well-defined or not, was not discussed. Manual testing still needs some means to identify the outcome of a test as correct or not, so there is no such thing as a manual test oracle in its own right.

Question 32   AI-3.2.3   K2

Which of the following sentences is NOT correct?

A. Classification and labeling algorithms are suited for a wide range of test data generation tasks.
B. A genetic algorithm can be used to generate structured input data.
C. Fuzz testing often is an appropriate method to generate test data.
D. A special form of neural networks can be used to generate images.

Justification

All three except A are given in section 3.2.3. The output of classification and labeling algorithms is usually not well-suited as test input.

Question 33   AI-3.4.3   K2

Match each item of the two lists to create correct statements?

1. An unsupervised learning method
2. Code coverage
3. A supervised learning method
4. Monkey testing

a. can be used for test selection.
b. requires the availability of past test performance.
c. cannot be used for test selection.
d. can be used as additional input.

A. 1a, 2b, 3d and 4c
B. 1a, 2c, 3d and 4b
C. 1b, 2c, 3a and 4d
D. 1a, 2d, 3b and 4c

Justification

The correct is answer D. As stated in section 3.4.3, an unsupervised learning method can always be used for test selection, even in the absence of labeled training data (i.e., past test performance). A supervised learning method on the other hand requires that kind of information. Code coverage can be used as additional input. But Monkey testing cannot be used for test selection.

Question 34   AI-3.3.2   K2

Why is test generation on the system level easier than on the component level?

1. The correct initial state
2. The input
3. An implicit test oracle
4. The output format

a. can be used.
b. can be much less formal.
c. is created by the system itself.
d. consists of primitive types.

A. 1a, 2b, 3c and 4d
B. 1c, 2d, 3a and 4b
C. 1b, 2c, 3a and 4d
D. 1a, 2d, 3b and 4c

Justification

The correct answer is B, according to 3.3.2: The correct initial state is created by the system itself, the input consists of primitive types, implicit test oracles can be used and the output format can be much less formal.

Question 35   AI-3.4.4   K2
What information can be used as input for an AI approach to identifier selection?
A.   Code coverage
B.   Screenshots of elements
C.   Confusion matrix
D.   Confidence in identifiers

Justification

A. Incorrect. Code coverage does not help with identifier selection.
B. Correct. The screenshot or "looks" of elements can be used as input to identifier selection in the case of image recognition, see section 3.4.4.
C. Incorrect. A confusion matrix is a table that summarizes how successful a classification model's predictions were.
D. Incorrect. The confidence in identifiers its output, not its input.

Question 36   AI-3.4.6   K2

What are the goals of visual test automation?

A. Identifying UI/UX problems
B. Identifying functional regressions
C. Identifying accessibility problems
D. Identifying visual regressions

Justification

A. Incorrect. As mentioned in section 3.2.2, AI cannot be used to suggest improvements to the user experience and user interface.
B. Incorrect. Visual test automation does not aim to test functional or business aspects of the software (see section 3.4.6).
C. Incorrect. Many accessibility aspects are not visual.
D. Correct. Visual test automation aims solely at identifying visual regressions or visual differences on different platforms (cross-platform and cross-device testing), see section 3.4.6.

Question 37   AI-3.5.1   K2

Which TWO of the following effects are emphasized by AI-based testing approaches?

A. Defect clustering
B. Pesticide paradox
C. Absence-of-errors fallacy
D. Self-fulfilling prophecy
E. No Free Lunch paradox

Justification

A. Incorrect. Defect clustering occurs on the code level and is not affected by testing techniques.
B. Correct. After some time, the pesticide paradox will then decrease the prediction efficiency (see section 3.5.1).
C. Incorrect. Absence-of-errors fallacy is not affected by AI techniques.
D. Correct. The self-fulfilling prophecy shows more errors in more intensely scrutinized areas of the code (see section 3.5.1).
E. Incorrect. The No Free Lunch theorem is not a paradox and not emphasized by AI-based testing.

Question 38   AI-3.4.2   K2

Which approaches can be used for test generation?

i.   Genetic algorithms
ii.   K-means clustering
iii.   Neural networks
iv.   Principal component analysis
v.   Support vector machines

A.   ii, iii true   i, iv, v false
B.   i, iii true   i, iv, v false
C.   i, iv true   ii, iii, v false
D.   ii, iv, v true   i, iii false

Justification

The correct answer is B. According to section 3.4.2, genetic algorithms and neural networks can be used for test generation, and even be combined. K-means clustering and principal component analysis as well as support-vector machines do not produce results applicable to test generation.

Question 39   AI-3.4.5   K2

Which TWO of the following AI approaches can be used for visual test automation?

A.   Image recognition
B.   Genetic algorithms
C.   Cross-platform testing
D.   Supervised learning algorithms
E.   Unsupervised learning algorithms

Justification

A.   Correct (see section 3.4.5).
B.   Incorrect. Genetic algorithms produce cannot be used to work with images.
C.   Incorrect. Cross-platform testing is not an AI approach.
D.   Correct (see section 3.4.5).
E.   Incorrect. Unsupervised learning algorithms cannot be used, as either elements or visual artifacts (defects) need to be identified, and thus labeled.

Question 40   AI-3.5.2   K3

Which of the following combination represents relevant aspects and questions to ask during a test generation tool evaluation?

1. Test case relevance
2. Defect relevance
3. Test case explosion
4. Maintainability

a. Which counter measures are in place?
b. How high is the expected customer impact?
c. Is there monitoring of effort and added benefit?
d. Are business use cases and associated risks covered?

A. 1d, 2b, 3a and 4c
B. 1a, 2c, 3d and 4b
C. 1b, 2c, 3a and 4d
D. 1c, 2d, 3b and 4a

Justification

The correct answer is A. Test cases should cover business use cases and associated risks (section 3.5.6). Revealed defects should have a high expected customer impact (section 3.5.9). There should be counter measures to test case explosion in place (at least a metric how many test cases are sensible, see section 3.5.7). And test case maintenance should be monitored in terms of effort and benefit (section 3.5.8).