

Sample Exam – Questions

Sample Exam set A
Version 2.0

ISTQB® AI Testing Syllabus

Compatible with Syllabus version 2.0

International Software Testing Qualifications Board



1 Copyright Notice

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

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

5 All rights reserved.

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

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

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

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

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

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

22 Document Responsibility

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

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

26 Acknowledgements

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

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

31

Revision History

32

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

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

33

Table of Contents

34		
35		
36	Copyright Notice	2
37	Document Responsibility.....	2
38	Acknowledgements.....	2
39	Revision History.....	3
40	Table of Contents.....	4
41	Introduction.....	6
42	Purpose of this document.....	6
43	Instructions.....	6
44	Questions	7
45	Question #1 (1 Point).....	7
46	Question #2 (1 Point).....	7
47	Question #3 (1 Point).....	8
48	Question #4 (1 Point).....	8
49	Question #5 (1 Point).....	8
50	Question #6 (1 Point).....	9
51	Question #7 (1 Point).....	9
52	Question #8 (1 Point).....	11
53	Question #9 (1 Point).....	11
54	Question #10 (1 Point).....	12
55	Question #11 (1 Point).....	13
56	Question #12 (1 Point).....	13
57	Question #13 (1 Point).....	14
58	Question #14 (1 Point).....	14
59	Question #15 (2 Points).....	14
60	Question #16 (1 Point).....	15
61	Question #17 (1 Point).....	15
62	Question #18 (1 Point).....	15
63	Question #19 (1 Point).....	16
64	Question #20 (1 Point).....	16
65	Question #21 (2 Points).....	16
66	Question #22 (1 Point).....	17
67	Question #23 (1 Point).....	17
68	Question #24 (1 Point).....	17
69	Question #25 (1 Point).....	18
70	Question #26 (1 Point).....	18
71	Question #27 (1 Point).....	18
72	Question #28 (2 Points).....	19
73	Question #29 (1 Point).....	19
74	Question #30 (1 Point).....	20
75	Question #31 (1 Point).....	20
76	Question #32 (1 Point).....	21
77	Question #33 (1 Point).....	21
78	Question #34 (2 Points).....	22
79	Question #35 (1 Point).....	23
80	Question #36 (1 Point).....	23
81	Question #37 (1 Point).....	24
82	Question #38 (1 Point).....	24
83	Question #39 (1 Point).....	24
84	Question #40 (1 Point).....	25
85	AppendiDx: Additional Questions.....	26
86	Question #A1 (1 Point).....	26



87	Question #A2 (1 Point)	26
88	Question #A3 (1 Point)	27
89	Question #A4 (1 Point)	27
90	Question #A5 (1 Point)	28
91	Question #A6 (1 Point)	28
92		

93 Introduction

94 Purpose of this document

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

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

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

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

103 Instructions

104 In this document you may find:

- 105 • Questions¹, including for each question:
 - 106 – Any scenario needed by the question stem
 - 107 – Point value
 - 108 – Response (answer) option set
- 109 • Additional questions, including for each question [does not apply to all sample exams]:
 - 110 – Any scenario needed by the question stem
 - 111 – Point value
 - 112 – Response (answer) option set
 - 113
- 114 • *Answers, including justification are contained in a separate document*

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

115

Questions

116

Question #1 (1 Point)

117

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

118

119

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

120

121

122

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

123

124

125

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

126

127

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

128

129

Select ONE answer.

130

131

Question #2 (1 Point)

132

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

133

134

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

135

136

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

137

138

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

139

140

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

141

142

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

143

144

Select TWO answers.

145

146 **Question #3 (1 Point)**

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

- 148 a) ML is a subset of AI, and DL is a further subset of ML, representing a hierarchy of
149 specialized technologies
- 150 b) DL is the foundational technology, and both ML and AI are specialized applications built
151 upon its principles.
- 152 c) DL and ML are essentially interchangeable terms, while AI represents a separate approach
153 to problem-solving
- 154 d) AI encompasses both ML and DL as distinct but important methodologies, working in
155 parallel to solve complex problems

156 Select ONE answer.

157

158 **Question #4 (1 Point)**

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

- 160 a) It is designed to analyze and understand existing data, such as text or images, allowing
161 machines to categorize information and extract key insights
- 162 b) It creates new content, such as text, images, or audio, by learning patterns from training
163 data and generating outputs similar in nature
- 164 c) It focuses mainly on improving existing content, such as text, images, or audio, by
165 optimizing it for classification or prediction tasks
- 166 d) It creates models that are limited to generating text and images and cannot be used for
167 creative tasks like drug discovery or data simulation

168 Select ONE answer.

169

170 **Question #5 (1 Point)**

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

- 173 a) GPUs are well-suited for training and running ML models due to their ability to handle
174 massively parallel processing, while CPUs are better for general-purpose computing
- 175 b) CPUs are more efficient than GPUs for ML tasks because they have faster clock speeds
176 and can handle complex operations
- 177 c) AI-specific hardware, such as ASICs, is primarily used for training ML models, while GPUs
178 are better suited for edge computing
- 179 d) Neuromorphic processors are a form of AI hardware specifically optimized for running on
180 the von Neumann architecture

181 Select ONE answer

182 **Question #6 (1 Point)**

183 Given the following statements about AI model development and hosting:

- 184 i. It achieves lower development costs by using public cloud resources, eliminating the need
185 for local hardware investment
- 186 ii. It uses local development of data preparation components for sensitive data for increased
187 security before moving to the cloud for training of the full system
- 188 iii. It results in lower costs because laptops are used for local development and there are low
189 upfront hardware costs by hosting AI models on public clouds
- 190 iv. It simplifies development and hosting by standardizing processes on local servers,
191 removing the need for complex cloud-based configurations
- 192 v. It guarantees the highest security by hosting AI models on private clouds, thereby avoiding
193 the risks associated with local hardware vulnerabilities

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

- 196 a) i, ii and v
- 197 b) ii and iii
- 198 c) iii and iv
- 199 d) i, iv and v

200 Select ONE answer.

201 **Question #7 (1 Point)**

202 Given the following ISO/IEC 25059 quality characteristics:

- 203 1. AI Robustness
- 204 2. User controllability
- 205 3. Functional adaptability
- 206 4. Intervenability

207 And the following examples of quality characteristic measures:

- 208 A. The success rate of a remote operator in forcing a drone into the safe-landing protocol
209 when its AI navigation system exhibits hazardous behavior
- 210 B. The average time required to successfully override a fraud management system's
211 automated decision to block a customer's transaction
- 212 C. The F1-score of an object detection model in an autonomous car in heavy rain
- 213 D. The time required for an e-commerce recommendation engine to update its suggestions to
214 reflect a new, rapidly emerging fashion trend

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

- 216 a) 1D – 2A – 3C – 4B
- 217 b) 1C – 2D – 3B – 4A
- 218 c) 1C – 2B – 3D – 4A
- 219 d) 1A – 2D – 3C – 4B

220 Select ONE answer.

221

222

223 **Question #8 (1 Point)**

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

- 226 a) When the requirements are too detailed, it leaves little room for the ML system to learn from
227 the implicit goals contained within the training data
- 228 b) The potential for non-determinism and self-learning makes some AI-based systems
229 unpredictable as they diverge from their original tested state
- 230 c) Because self-learning safety-related systems stop adapting after deployment, safety is
231 often compromised by the need for manual updates of the operational AI-based system
- 232 d) The mature safety-related standards tend to be out-of-date and require the use of outdated
233 AI technologies, which hinders innovative AI solutions in AI-based systems

234 Select ONE answer.

235 **Question #9 (1 Point)**

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

- 238 a) The security guard in the museum control room can trigger an immediate 'all-stop'
239 command that causes the patrol robot to cease all motion within 0.5 seconds to prevent
240 collision with a sculpture
- 241 b) A greenhouse control system reacts within 20 minutes when the measured humidity is
242 greater than 10% from the optimum humidity
- 243 c) The spam alert control system is easy for the user to set up and requires minimal technical
244 expertise to maintain
- 245 d) When the analysis tool flags a retinal scan for severe diabetic retinopathy, it shall display a
246 visual heatmap overlay on the image, highlighting the key features

247 Select ONE answer.

248

249 **Question #10 (1 Point)**

250 Given the following forms of ML:

- 251 1. Clustering
252 2. Reinforcement learning
253 3. Classification
254 4. Regression

255 And the following examples:

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

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

- 265 a) 1B – 2C – 3D – 4A
266 b) 1A – 2D – 3B – 4C
267 c) 1C – 2D – 3A – 4B
268 d) 1D – 2C – 3B – 4A

269 Select ONE answer.

270

271 **Question #11 (1 Point)**

272 Given the following activities from the ML workflow:

- 273 1. Deploy the Model
274 2. Prepare & Test Data
275 3. Test the Model
276 4. Evaluate the Model

277 And, given the following descriptions:

- 278 A. Model performance is tested using validation data
279 B. The origin of the test data used to test the model is identified
280 C. Test data are used to verify the agreed performance criteria are met
281 D. The model is tested on the target platform
282

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

- 284 a) 1A – 2C – 3B – 4D
285 b) 1C – 2D – 3B – 4A
286 c) 1D – 2B – 3A – 4C
287 d) 1D – 2B – 3C – 4A

288 Select ONE answer.

289 **Question #12 (1 Point)**

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

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

299 Select ONE answer.

300

301 **Question #13 (1 Point)**

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

- 304 a) Data preparation comprises the identification and gathering of data from various sources,
305 providing the algorithm with raw data to learn from
- 306 b) Feature engineering of the data is performed after an ML model has been trained to
307 optimize its performance for real-world deployment
- 308 c) Data pre-processing includes augmentation and sampling, which either add or reduce the
309 number of examples in the training data respectively
- 310 d) Exploratory data analysis (EDA) is a form of exploratory testing applied to the data
311 preparation activities of acquisition, pre-processing and labelling

312 Select ONE answer.

313 **Question #14 (1 Point)**

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

- 316 a) The training dataset is used to optimize hyperparameters, the validation dataset is used to
317 tune predictions, and the test dataset is used to generate training data
- 318 b) The training dataset is used to create the model, the validation dataset is used to tune the
319 model, and the test dataset evaluates its performance on unseen data
- 320 c) The training dataset is used for final model evaluation, the validation dataset ensures the
321 model does not overfit, and the test dataset is used for tuning hyperparameters
- 322 d) The training dataset ensures the model generalizes well, the validation dataset is used to
323 deploy the model, and the test dataset is used for initial evaluation

324 Select ONE answer.

325 **Question #15 (2 Points)**

326 Consider the following confusion matrix for an image classifier:

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

327
328

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

- 331
- 332 a) $(20/120) * 100$
- 333 b) $(78/120) * 100$
- 334 c) $(78/100) * 100$
- 335 d) $(22/100) * 100$

336 Select ONE answer.

337 **Question #16 (1 Point)**

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

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

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

345 Select ONE answer.

346 **Question #17 (1 Point)**

347 Given the following examples of AI-based systems:

- 348 i. A system that learns from real-time data to improve its failure predictions and automatically
349 updates maintenance schedules
350 ii. A spam filter for an email app, which identifies spam based on predefined rules
351 iii. A recommendation engine on a streaming service that updates its suggestions based on a
352 user's changing viewing habits and preferences
353 iv. A personal assistant that learns continuously from its user
354 v. A rule-based system for medical diagnosis
355

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

- 358 a) ii and v
359 b) i, iii, and iv
360 c) i, and iii
361 d) ii, iv, and v
362

363 Select ONE answer.

364 **Question #18 (1 Point)**

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

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

375 Select ONE answer.

376

377 **Question #19 (1 Point)**

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

- 380 a) It is a method that only functions for AI systems built from complete and detailed technical
381 specifications
- 382 b) It ensures reproducibility for individual test runs but cannot alter the model's inherent
383 probabilistic nature in production
- 384 c) Its success depends upon extensive consultation with domain experts to select the most
385 appropriate seed value for the tests
- 386 d) It introduces a high degree of subjectivity into the model's behavior, which complicates the
387 test evaluation process.

388 Select ONE answer.

389 **Question #20 (1 Point)**

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

- 391 a) GenAI models are tested by verifying that their outputs exactly match predefined expected
392 results
- 393 b) It involves manipulating diverse inputs and parameters and then assessing the output's
394 adherence to rules, because a direct input-to-output match is not feasible
- 395 c) Since GenAI models are probabilistic, formal testing is unnecessary because outputs will
396 always vary and cannot be evaluated effectively
- 397 d) GenAI models are mainly tested through manual review, as automated testing does not apply
398 to creative AI outputs

399 Select ONE answer.

400 **Question #21 (2 Points)**

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

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

- 408 a) Prioritize security vulnerabilities first, then address bias issues in a separate phase after
409 deployment
- 410 b) Focus on bias detection since legal advice accuracy is more critical than security concerns
- 411 c) Deploy immediately and conduct red teaming reactively based on incidents and user
412 feedback
- 413 d) Use attack scenarios covering both security and bias vulnerabilities before deployment

414 Select ONE answer.

415 **Question #22 (1 Point)**

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

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

- 420 a) System testing
421 b) Input data testing
422 c) Component integration testing
423 d) ML model testing

424 Select ONE answer.

425 **Question #23 (1 Point)**

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

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

435 Select ONE answer.

436 **Question #24 (1 Point)**

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

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

- 441 a) Data provenance testing
442 b) Data representativeness testing
443 c) Feature testing
444 d) Dataset constraint testing

445 Select ONE answer.
446

447

448 **Question #25 (1 Point)**

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

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

- 454 a) Conducting static analysis of the model's source code to identify how its operations on the age and
455 gender related attributes could lead to bias
- 456 b) Testing using a dataset that is representative and analyzing the predictions for statistically significant
457 differences in outcomes across age and gender
- 458 c) Reviewing the overall ML workflow and data preparation processes to identify potential sources of
459 bias introduction
- 460 d) Performing disparate impact analysis using counterfactuals based on gender, age or both

461 Select ONE answer.

462 **Question #26 (1 Point)**

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

- 465 a) Testing an operational pipeline would focus mainly on validating individual transformation scripts,
466 whereas a training pipeline's testing would prioritize end-to-end system testing
- 467 b) A training pipeline would rely almost exclusively on fault injection and back-to-back tests, while an
468 operational pipeline would be limited to unit and component integration testing
- 469 c) Configuration management reviews would be critical for operational pipelines but are considered
470 unnecessary for the less formal nature of exploratory training pipelines
- 471 d) Testing a training pipeline would primarily focus on determining that data is handled correctly, while
472 an operational pipeline's testing would emphasize high performance and reliability under load

473 Select ONE answer.

474 **Question #27 (1 Point)**

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

- 476 a) It serves as a universally applicable dataset derived from trusted industry benchmarks
- 477 b) It functions as a quantitative baseline for formally comparing the statistical properties of the
478 training data
- 479 c) It is used to apply stratified sampling directly to the training data to verify that all subgroups
480 are covered
- 481 d) It serves as the primary dataset for performing the final model validation and testing with
482 high independence

483 Select ONE answer.

484

485 **Question #28 (2 Points)**

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

- 488 • applicant_id (integer, unique)
- 489 • annual_income (decimal, in USD)
- 490 • loan_amount (decimal, in USD)
- 491 • credit_score (integer, 300-850)
- 492 • employment_years (integer, 0-50)
- 493 • monthly_payment (decimal, in USD)

494 The following business rules apply:

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

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

- 501 a) Single-value range constraint
- 502 b) Multi-value count constraint
- 503 c) Comparison correlate constraint,
- 504 d) Multi-value duplicate constraint

505 Select ONE answer.

506 **Question #29 (1 Point)**

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

- 509 a) Multiple annotations can be used to automate label verification to ensure consistency
- 510 b) Multiple annotations can point to label defects based on the comparison of annotations
- 511 c) Multiple annotations compare label distributions across datasets to detect potential defects
- 512 d) Multiple annotations point to defects in data points with high model loss

513 Select ONE answer.

514

515 **Question #30 (1 Point)**

516 Given the following test techniques and test types:

- 517 1. ML functional performance testing
518 2. Testing for bias
519 3. Adversarial testing
520 4. Drift testing
521

522 And the following risks:

- 523 A. The ML model might perform differently for different demographic groups
524 B. Slightly modified inputs to the ML model might cause quite different and unexpected
525 responses
526 C. Predictions made by the ML model might be inaccurate in some cases
527 D. ML model accuracy might have significantly decreased since it was deployed
528

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

- 530 a) 1A – 2D – 3C – 4B
531 b) 1C – 2A – 3B – 4D
532 c) 1B – 2A – 3C – 4D
533 d) 1A – 2C – 3D – 4B
534

535 Select ONE answer.

536 **Question #31 (1 Point)**

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

- 538
539 a) Information related to the speed of model prediction is a part of the documentation of an AI
540 component
541 b) Interaction of AI and non-AI components is within the scope of the model documentation
542 c) Bias-related characteristics of input data are challenging to assess as the model
543 documentation doesn't contain the source of the training data
544 d) Changes made by self-learning ML systems are fully documented at the time of the update
545 to ensure up-to-date documentation

546 Select ONE answer.

547

548 **Question #32 (1 Point)**

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

551
552 Previously, the ML model had achieved an accuracy of $83\% \pm 4\%$ at a confidence level of 94%.

553
554 Which of the following options is **MOST** likely to represent the current situation?

- 555 a) accuracy of $83\% \pm 2\%$ at 94% confidence level
556 b) accuracy of $83\% \pm 4\%$ at 92% confidence level
557 c) accuracy of $83\% \pm 6\%$ at 94% confidence level
558 d) accuracy of $85\% \pm 4\%$ at 94% confidence level

559 Select ONE answer.

560 **Question #33 (1 Point)**

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

- 563 a) It is a form of black-box testing, ignoring knowledge of the internals of the machine learning
564 system to create adversarial examples
565 b) It focuses on generating manual adversarial examples, without automation, to test the
566 vulnerability of machine learning systems
567 c) It identifies model vulnerabilities through adversarial examples, which are minimally
568 perturbed inputs that induce misclassification
569 d) It verifies the machine learning system's functionality by using previously working tests to
570 avoid ML model failures during evaluation and tuning

571 Select ONE answer.

572

573 **Question #34 (2 Points)**

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

577
578 Metamorphic testing is being used with the following source test case:

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

Outputs
Recommended Phones: SnapHappy_X1 SnapHappy_M2 SnapHappy_M3 ClickNow_1000x ClickNow_1000xs

579
580
581 And this test data for two corresponding follow-up test cases:

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

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

582
583 Which of the following options is **MOST** likely to be a valid list of recommended phones for the
584 follow-up test cases?

- 585 a) T1: SnapHappy_X1, SnapHappy_M2
586 T2: ClickNow_1000x, ClickNow_1000xs
- 587 b) T1: SnapHappy_M2, SnapHappy_M3, ClickNow_1000xs
588 T2: SnapHappy_X1, ClickNow_1000x
- 589 c) T1: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3, ClickNow_1000x,
590 ClickNow_1000xs
591 T2: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3
- 592 d) T1: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3, ClickNow_1000x,
593 ClickNow_1000xs
594 T2: SnapHappy_X1, SnapHappy_M2, SnapHappy_M3, ClickNow_1000x,
595 ClickNow_1000xs
596

597 Select ONE answer.

598 Question #35 (1 Point)

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

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

- 603 a) Dynamic drift testing, because it can infer the current ground truth by analyzing the input
604 data's statistical properties
- 605 b) Static drift testing, because it identifies drift by detecting changes in the data distributions
606 without requiring ground truth
- 607 c) Dynamic drift testing, because comparing ML model predictions to actual results is the most
608 direct way to measure performance degradation
- 609 d) Static drift testing, because it compares the ML model's live performance metrics against a
610 predefined acceptance threshold

611 Select ONE answer.

612 Question #36 (1 Point)

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

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

- 616 a) Underfitting
- 617 b) Concept drift
- 618 c) Overfitting
- 619 d) Low acceptance criteria

620 Select ONE answer.

621

622 **Question #37 (1 Point)**

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

- 625 a) A/B testing is primarily used to generate diverse test cases that cover all possible inputs for
626 an MLS
- 627 b) A/B testing is used to verify that all components within an MLS interact correctly with each
628 other
- 629 c) A/B testing determines if an updated version of an MLS performs better than the previous
630 version
- 631 d) A/B testing focuses on analyzing the internal algorithm structure of an MLS to identify
632 potential defects

633 Select ONE answer.

634 **Question #38 (1 Point)**

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

- 637 a) By varying the hyperparameters used for training the ML model that is being tested
- 638 b) By fine-tuning the ML model that is being tested
- 639 c) By supplementing the ML model that is being tested with retrieval-augmented generation
- 640 d) By using a different ML development framework than the ML model that is being tested

641 Select ONE answer.

642 **Question #39 (1 Point)**

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

- 645 a) A team notices that their ML development framework becomes slow and even
646 unresponsive when processing large batches of data
- 647 b) A new version of a core library used by the ML development framework appears to be
648 causing the ML model to produce unexpected and inaccurate results
- 649 c) After a new installation, the development team needs a quick test to determine if the ML
650 development framework's essential services are running correctly
- 651 d) A single test result is difficult to reproduce during evaluation, showing slight variations in ML
652 functional performance with each execution
- 653 e) A project leader needs to decide between two potential algorithms based on their suitability
654 for the project's goals

655 Select TWO answers.

656

657 **Question #40 (1 Point)**

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

- 660 a) Shadow testing affects the responses from real users, while canary testing does not involve
661 real users
- 662 b) Shadow testing mirrors traffic without impacting users, while canary testing provides
663 responses from the new ML model
- 664 c) Canary testing is used for performance testing, while shadow testing is mainly used for
665 component integration testing
- 666 d) Canary testing compares ML models running offline, while shadow testing is based on the
667 use of live user data

668 Select ONE answer.

669

670

AppendiDx: Additional Questions

671

Question #A1 (1 Point)

672

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

An artificial mind that generates entirely new forms of art, music, and mathematics that are incomprehensible to humans.	Narrow AI
A system that manages complex daily schedules, learns new recipes from a video, and holds conversations about novels it has just read.	General AI
A system that can independently learn any field of science and collaborate with human scientists by proposing novel hypotheses and original experiments.	Super AI
A system that examines radiological images to detect the specific signatures of cancerous tumors.	
A language translation model that can convert written text from French to Spanish.	

673

674

675

676

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

677

Question #A2 (1 Point)

678

679

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

680

681

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

682

683

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

684

685

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

686

687

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

688

Select ONE answer.

689

690 **Question #A3 (1 Point)**

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

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

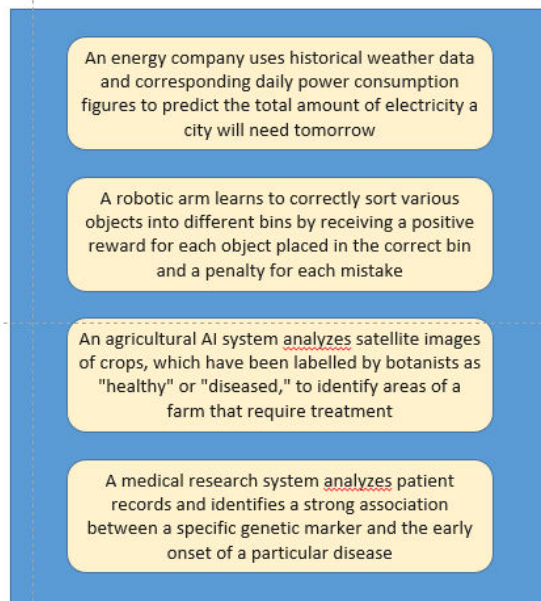
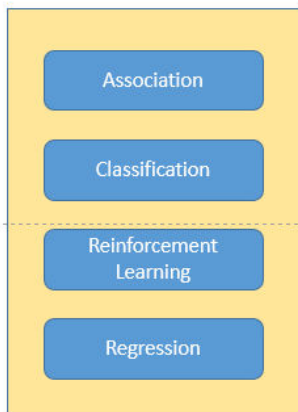
- 695 a) ISO/IEC/IEEE TR 29119-11
- 696 b) OECD AI Principles
- 697 c) EU AI Act
- 698 d) ISO/IEC 25059

699 Select ONE answer.

700

701 **Question #A4 (1 Point)**

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



703

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

705

706 **Question #A5 (1 Point)**

707 Given the following activities:



708

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

710 **Question #A6 (1 Point)**

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

- 713 a) It tends to overestimate performance on tasks involving human feedback
- 714 b) It cannot account for differences in model size when comparing test results
- 715 c) It does not reveal whether the network relies on misleading or irrelevant features
- 716 d) It replaces the need to assess the usefulness of individual model outputs

717 Select ONE answer.