# Sample Exam – Answers

**Sample Exam set A**
**Version 1.1**

# ISTQB® Testing with Generative AI Syllabus
# Specialist Level

**Compatible with Syllabus version 1.1**

International Software Testing Qualifications Board


Version 1.1 27/04/2026

# Copyright Notice



# Document Responsibility

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

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

# Acknowledgements

This document was produced by a core team from the ISTQB®: Abbas Ahmad, Alessandro Collino, Bruno Legeard.

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

# Revision History

| Sample Exam – Answers Layout Template used: | Version 2.12 Date: November 27, 2024 |
|---|---|

| Version | Date | Remarks |
|---|---|---|
| V1.0 | 2024/12/30 | CT-GenAI Sample Exam set A 1.0 Alpha Release |
| V1.0 | 2025/04/17 | CT-GenAI Sample Exam set A 1.0 Beta Release |
| V1.0 | 2025/06/10 | CT-GenAI Sample Exam set A 1.0 Release |
| V1.1 | 2026/04/27 | CT-GenAI Sample Exam set A 1.1 Minor Update |

# Table of Contents

# Introduction

## Purpose of this document

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

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

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

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

## Instructions

In this document you may find:

- 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 | d | GenAI-1.1.1 | K1 | 1 |
| 2 | c | GenAI-1.1.2 | K2 | 1 |
| 3 | b | GenAI-1.1.2 | K2 | 1 |
| 4 | d | GenAI-1.1.3 | K2 | 1 |
| 5 | b | GenAI-1.1.4 | K2 | 1 |
| 6 | a, e | GenAI-1.2.1 | K2 | 1 |
| 7 | d | GenAI-1.2.2 | K2 | 1 |
| 8 | b | GenAI-2.1.1 | K2 | 1 |
| 9 | c | GenAI-2.1.1 | K2 | 1 |
| 10 | b | GenAI-2.1.2 | K2 | 1 |
| 11 | a | GenAI-2.1.3 | K2 | 1 |
| 12 | a | GenAI-2.2.1 | K3 | 2 |
| 13 | b | GenAI-2.2.2 | K3 | 2 |
| 14 | c | GenAI-2.2.3 | K3 | 2 |
| 15 | c | GenAI-2.2.4 | K3 | 2 |
| 16 | a | GenAI-2.2.5 | K3 | 2 |
| 17 | a | GenAI-2.3.1 | K2 | 1 |
| 18 | a | GenAI-2.3.2 | K2 | 1 |
| 19 | c | GenAI-3.1.1 | K1 | 1 |
| 20 | d | GenAI-3.1.2 | K3 | 2 |

| Question Number (#) | Correct Answer | LO | K-Level | Points |
|---|---|---|---|---|
| 21 | b | GenAI-3.1.3 | K2 | 1 |
| 22 | b | GenAI-3.1.4 | K1 | 1 |
| 23 | d | GenAI-3.2.1 | K2 | 1 |
| 24 | d | GenAI-3.2.2 | K2 | 1 |
| 25 | a | GenAI-3.2.2 | K2 | 1 |
| 26 | b | GenAI-3.2.3 | K2 | 1 |
| 27 | c | GenAI-3.3.1 | K2 | 1 |
| 28 | b, d | GenAI-3.4.1 | K1 | 1 |
| 29 | a | GenAI-4.1.1 | K2 | 1 |
| 30 | a | GenAI-4.1.2 | K2 | 1 |
| 31 | c | GenAI-4.1.3 | K2 | 1 |
| 32 | b | GenAI-4.2.1 | K2 | 1 |
| 33 | b | GenAI-4.2.2 | K2 | 1 |
| 34 | d | GenAI-5.1.1 | K1 | 1 |
| 35 | b | GenAI-5.1.2 | K2 | 1 |
| 36 | b | GenAI-5.1.3 | K2 | 1 |
| 37 | a | GenAI-5.1.4 | K1 | 1 |
| 38 | c | GenAI-5.2.1 | K2 | 1 |
| 39 | c | GenAI-5.2.2 | K1 | 1 |
| 40 | a | GenAI-5.2.3 | K1 | 1 |

# Answers

| Question Number (#) | Correct Answer | Explanation / Rationale | Learning Objective (LO) | K-Level | Number of Points |
|---|---|---|---|---|---|
| **1** | d | Considering:<br>• Symbolic AI uses a rule-based system to mimic human decision-making, representing knowledge using symbols and logical rules (1B).<br>• Classical machine learning uses a data-driven approach that requires data preparation, feature selection and model training (2D).<br>• Deep learning uses neural networks (machine learning structures) to automatically learn features from data (3A).<br>• Generative AI uses deep learning techniques to create new data by learning and mimicking patterns from its training data (4C).<br><br>Thus:<br>a) Is not correct.<br>b) Is not correct.<br>c) Is not correct.<br>d) Is correct. | GenAI-1.1.1 | K1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **2** | c | a) Is not correct because context windows do not control temporal sequencing but limit the scope of text that can be simultaneously considered.<br>b) Is not correct because context windows affect scope within the current input, not cross-document referencing capabilities.<br>c) Is correct. When text exceeds the context window size, the model cannot simultaneously consider all parts of the document. As the model processes new tokens, it must effectively "forget" or discard tokens that fall outside its context window boundary.<br>d) Is not correct because context windows do not determine parsing approaches but limit the amount of text that can be processed simultaneously. | GenAI-1.1.2 | K2 | 1 |
| **3** | b | a) Is not correct. This describes embeddings, not tokenization.<br>b) Is correct. Tokenization involves splitting text into smaller units (tokens) that represent the building blocks of natural language generation tasks and enable LLMs to understand and generate text. See also the definition of "Tokenization" in the syllabus (see "Appendix D – Generative AI Specific Terms").<br>c) Is not correct. This describes the general function of LLMs, not tokenization.<br>d) Is not correct. This refers to how LLMs generate text, not tokenization. | GenAI-1.1.2 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **4** | d | Considering:<br>i. Is not correct. While foundation LLMs can generate test cases, they do not inherently "excel" at this task without structured input. Generating test cases without structured input misrepresents their capabilities.<br>ii. Is not correct. Creating template-based test scripts requires executing explicit instructions, which is the role of instruction-tuned LLMs, not reasoning LLMs. Reasoning LLMs specialize in logical inference and problem-solving, not rigid template adherence.<br>iii. Is not correct. Instruction-tuned LLMs are designed to follow structured prompts, not make autonomous decisions. Prioritizing tests based on feedback requires reasoning, which is beyond their scope.<br>iv. Is correct. Reasoning LLMs are explicitly designed for synthesizing multiple data sources and performing logical inference, problem-solving, and decision-making. Detecting trends and prioritizing test efforts aligns with their role in contextual analysis and the description provided.<br>v. Is correct. Instruction-tuned LLMs are specifically trained to follow instructions, including adhering to requested formats, styles, and syntax rules. Generating test cases conforming to Gherkin language syntax, is a task well-suited to their capabilities.<br><br>Thus:<br>a) Is not correct.<br>b) Is not correct.<br>c) Is not correct.<br>d) Is correct. | GenAI-1.1.3 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **5** | b | a) Is not correct. Vision-language models are a subset of multimodal LLMs, not the reverse.<br>b) Is correct. Vision-language models specifically integrate visual and textual data, making them a subset of multimodal LLMs.<br>c) Is not correct. Vision-language models are closely related to multimodal LLMs and focus on both visual and textual data.<br>d) Is not correct. Multimodal LLMs and vision-language models have distinct scopes and are not interchangeable. | GenAI-1.1.4 | K2 | 1 |
| **6** | a, e | a) Is correct. LLMs can analyze and clarify requirements by identifying ambiguities and inconsistencies.<br>b) Is not correct. Generating complete application code is not a key capability of LLMs in test tasks.<br>c) Is not correct. LLMs can support test automation by suggesting improvements to test scripts and identifying design patterns but do not directly execute test scripts or fully automate test scripts without human oversight.<br>d) Is not correct. LLMs cannot perform manual exploratory testing because it is an intuitive and adaptive process that requires human creativity, experience, and decision-making.<br>e) Is correct. LLMs can generate diverse test data, including combinations and boundaries. | GenAI-1.2.1 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **7** | d | a) Is not correct. AI chatbots are best suited for ad hoc interactions, not specific test tasks.<br>b) Is not correct. AI chatbots and LLM-powered testing applications have distinct purposes and are not identical in functionality or configurability.<br>c) Is not correct. LLM-powered testing applications focus on integration into test processes, not conversational prompts. AI chatbots do not require integration into test tools and into test processes because their primary function is to facilitate ad hoc interactions rather than performing specific test tasks.<br>d) Is correct. AI chatbots provide conversational interfaces for ad hoc tasks, while LLM-powered testing applications deliver customized solutions for specific needs. | GenAI-1.2.2 | K2 | 1 |
| **8** | b | a) Is not correct. Context provides background information about the test environment or system being tested, not specific (input) data to be analyzed. While this mentions performance testing, it is listing actual data sources rather than providing background information.<br>b) Is correct. The description lists specific data sources (test reports, monitoring logs, performance benchmarks) that will be processed by the LLM to perform the analysis task.<br>c) Is not correct. Constraints outline restrictions or special considerations for how the task should be performed. The description lists actual data sources rather than restrictions on the analysis approach.<br>d) Is not correct. Output format specifies the expected structure and characteristics of the LLM's response. The description lists actual data sources, without any reference to how results should be presented. | GenAI-2.1.1 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **9** | c | a) Is not correct. The given line does not describe what task to perform, but rather how to present the results. Instructions tell the LLM what to do, while this line tells the LLM how to format the output.<br>b) Is not correct. The given line is not imposing restrictions or limitations on the analysis process, but rather specifying the desired presentation format. Constraints might include aspects like “exclude cosmetic issues” whereas this line is about formatting the output.<br>c) Is correct. The given line matches the syllabus definition of output format specifications that guide how the LLM should format the output.<br>d) Is not correct. The given line provides no background information such as the context of the requirements specification. Instead, it is about formatting the output. | GenAI-2.1.1 | K2 | 1 |
| **10** | b | a) Is not correct. Prompt chaining is actually defined by decomposing a task into sequential prompts, not by “providing examples”. Few-shot prompting is the technique that supplies examples, not the one that “breaks tasks into subtasks”. Meta prompting relies on the LLM to revise prompts by itself, not on the tester “refining prompts manually”.<br>b) Is correct. Few-shot prompting provides guidance with examples, prompt chaining decomposes tasks into intermediate steps (multiple prompts), and meta prompting uses AI to refine its own prompts iteratively.<br>c) Is not correct. Meta prompting’s core purpose is automatic prompt refinement by the LLM, not “breaking tasks into steps”. Prompt chaining is characterized by step-by-step decomposition, not by “using examples” as stated. Few-shot prompting is the method that supplies examples to the model and it is not centered on “manual optimization of prompts”.<br>d) Is not correct. Prompt chaining is defined by stepwise decomposition, not “guidance without examples”. Meta prompting relies on the LLM to generate/refine prompts, not solely on tester-defined wording. | GenAI-2.1.2 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **11** | a | a) Is correct. The system prompt stays constant throughout the interaction session and establishes the fundamental framework for how the LLM should respond.<br>b) Is not correct. This describes the function of a user prompt, not a system prompt.<br>c) Is not correct. The system prompt does not adjust dynamically; it remains constant.<br>d) Is not correct. The system prompt is hidden and does not include visible input from the user. | GenAI-2.1.3 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **12** | a | Considering:<br>i. Is correct because it ensures the process begins with generating test conditions from the requirements which aligns with the syllabus description that test analysis with GenAI involves generating test conditions based on the test basis, for example on requirements.<br>ii. Is correct because it clarifies the need for acceptance criteria (test conditions) to improve LLM-generated outputs and follows the syllabus guidance that LLMs can prioritize test conditions based on risk level when provided with proper context.<br>iii. Is correct because it directs the LLM to perform coverage analysis, addressing all aspects of the requirements, which matches the syllabus statement that LLMs can perform coverage analysis to determine whether all aspects of the test basis are covered.<br>iv. Is not correct because relying on minimal input without guidance does not align with the prompt chaining technique or ensure effectiveness. This options attempts to do everything in one step rather than breaking it down.<br>v. Is not correct because defect identification is not central to the test objective of generating prioritized test conditions and identifying coverage gaps. The scenario specifically states the requirements are stable and already thoroughly reviewed for defects (which typically include ambiguities and inconsistencies).<br><br>Thus:<br>a) Is correct.<br>b) Is not correct.<br>c) Is not correct.<br>d) Is not correct. | GenAI-2.2.1 | K3 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| **13** | b | a) Is not correct. While it specifies using predefined examples, it does not explicitly require the use of the "Given-When-Then" syntax, which is crucial for Gherkin-style test cases. It also indicates reliance on vague best practices in defining the constraints and does not specifically ensure alignment with the acceptance criterion.<br>b) Is correct. Comprehensive and leverages predefined examples to guide the LLM.<br>c) Is not correct. Lacks emphasis on using predefined examples and indicates reliance on vague best practices in defining the instructions.<br>d) Is not correct. Focuses on edge cases but neglects comprehensive coverage and the use of examples for guidance. | GenAI-2.2.2 | K3 | 2 |
| **14** | c | a) Is not correct. Addresses clustering and cross-checking but misses other critical steps like separating test results.<br>b) Is not correct. Improves role clarity but does not expand instructions or address structured steps.<br>c) Is correct. Expands instructions to include all critical structured analysis steps. Separating expected results and actual results helps to pinpoint mismatches effectively, clustering issues facilitates better prioritization and reduces redundancy, and highlighting discrepancies helps to focus attention on the most important findings.<br>d) Is not correct. Introduces irrelevant constraints, misaligning the prompt with the task requirements. | GenAI-2.2.3 | K3 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| **15** | c | a) Is not correct. Clarifies the role but does not add any concrete instruction that improves accuracy, actionability, or interpretability of the delivered metrics.<br>b) Is not correct. Adds a risk-analysis task that could distract the model from the core metrics and still offers no mechanism for making results clear to stakeholders.<br>c) Is correct. Expanding the output format with a plain-language summary that interprets the metrics and outlines next steps directly supports stakeholder understanding and actionability (see also the syllabus in section 2.2.4: “Enhanced test metrics visualization and reporting”).<br>d) Is not correct. Merely re-states an existing constraint and gives no specific guidance on how the LLM should achieve stakeholder-level interpretability. | GenAI-2.2.4 | K3 | 2 |
| **16** | a | a) Is correct. Few-shot prompting is ideal in this scenario because it allows you to provide some examples (the existing test cases with inputs and expected results) to guide the LLM. By illustrating how these transformation rules are applied to generate new test cases, few-shot prompting can help the LLM understand and replicate the process to generate additional test cases.<br>b) Is not correct. While prompt chaining could be used, it could unnecessarily complicate this very straightforward task.<br>c) Is not correct. While meta-prompting could be used, it is not as direct and specific as few-shot prompting to address this very straightforward task .<br>d) Is not correct. Zero-shot prompting would not be effective since it would not leverage existing test cases as examples. | GenAI-2.2.5 | K3 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| **17** | a | a) Is correct. Diversity ensures comprehensive coverage of edge cases, and test execution success rate evaluates the reliability of API test scripts (see the description and example for the "diversity" metric provided by the syllabus within the table in section 2.3.1).<br>b) Is not correct. While accuracy and completeness are important, relying primarily on time efficiency does not fully evaluate coverage or test execution reliability.<br>c) Is not correct. Precision and contextual fit are valuable but do not address diversity or thoroughly evaluate test execution success for API test scripts.<br>d) Is not correct. Relevance and contextual fit are crucial, but without considering metrics like test execution success rate or accuracy, critical aspects of evaluation are missed. | GenAI-2.3.1 | K2 | 1 |
| **18** | a | a) Is correct. Output analysis examines LLM outputs for inaccuracies or inconsistencies. By classifying the wrong expected results that contradict the requirements, it reveals why the prompt misled the LLM and provides concrete insights for prompt refinement.<br>b) Is not correct. A/B testing of prompts is better suited for comparing different prompt versions than for diagnosing the cause of wrong expected results.<br>c) Is not correct. While adjusting the length and specificity of prompts can improve the quality of responses by adding or reducing context, it does not directly address the cause of wrong expected results.<br>d) Is not correct. While using the insights gathered from testers about the usefulness and clarity of generated output can help refine prompts to better meet real-world testing needs, it does not directly address the cause of wrong expected results. | GenAI-2.3.2 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **19** | c | a) Is not correct. This describes reasoning errors, not hallucinations.<br>b) Is not correct. This described biases in AI output, not hallucinations.<br>c) Is correct. Hallucinations occur when the LLM generates output that is factually incorrect or irrelevant to a given task. See the syllabus in section 3.1.1.<br>d) Is not correct. This describes biases due to underrepresentation in training data, not hallucinations. | GenAI-3.1.1 | K1 | 1 |
| **20** | d | a) Is not correct. Cart management is explicitly mentioned in the project briefing, making this test case relevant.<br>b) Is not correct. Discount code application is explicitly mentioned in the project briefing, making this test case relevant.<br>c) Is not correct. Order confirmation emails are explicitly mentioned in the project briefing, so this test case is valid.<br>d) Is correct. Wishlist management is not mentioned in the project briefing, making this the most likely hallucinated test case. | GenAI-3.1.2 | K3 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| **21** | b | a) Is not correct. The effort to fine-tune LLMs for test tasks is associated with further training of these models on a targeted dataset, enabling them to learn the domain-specific knowledge and nuances needed to perform those test tasks. The use of clear and structured input data formats does not affect this effort.<br>b) Is correct. When input data is presented in a clear and structured way, potential misunderstandings are minimized. Ambiguity often arises from unclear and poorly structured data.<br>c) Is not correct. Context relevance is more dependent on providing the right contextual information rather than just using clear and structured input data formats.<br>d) Is not correct. The use of clear and structured input data formats does not increase the creativity of LLMs in generating outputs. In particular, the use of these formats does not encourage novel responses, as it requires LLMs to generate responses that adhere to the specified formats. | GenAI-3.1.3 | K2 | 1 |
| **22** | b | a) Is not correct. Learning rate is a setting (chosen before training begins) that controls the learning process of neural networks. This setting determines how much to change the model's weights during training. It is not related to inference variability. Increasing it can lead to faster convergence during training but does not affect the variability of outputs during inference.<br>b) Is correct. Temperature is a parameter that controls the randomness of the output by acting on the probability distribution during inference. Lowering the temperature reduces randomness, leading to more consistent outputs.<br>c) and d) are not correct. While setting a random seed can improve reproducibility, it does not in itself narrow the probability distribution during inference. Moreover, whether its value is high or low is irrelevant. | GenAI-3.1.4 | K1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **23** | d | a) Is not correct. This is a privacy concern, as “Unintentional data exposure” relates to generative AI (GenAI) outputs revealing sensitive information accidentally.<br>b) Is not correct. This is a privacy concern, as it involves the lack of control over how sensitive data is stored or processed.<br>c) Is not correct. This is a privacy concern related to non-compliance with regulations such as the General Data Protection Regulation (GDPR), leading to legal risks.<br>d) Is correct. An LLM does not expose real sensitive data if it hallucinates when generating synthetic test data, as long as it was not trained on real sensitive data. In this case hallucinations would be purely synthetic and based on patterns and structures the model learned during training. The LLM could unintentionally generate synthetic test data that matches real sensitive data (and this is definitely a concern) but it would not be exposing real sensitive data. However, this unintentional generation is very unlikely. | GenAI-3.2.1 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **24** | d | a) Is not correct. Malicious code generation involves “manipulating an LLM to generate backdoors (e.g., external command calls) during use”, not injecting false data into training datasets.<br>b) Is not correct. Context Manipulation involves “sending requests designed to extract confidential training data”, not injecting false data into training datasets.<br>c) Is not correct. Request manipulation involves “introducing data that disrupts the AI’s output” during runtime use, not injecting false data into training datasets.<br>d) Is correct. According to the syllabus in section 3.2.2, data poisoning involves “manipulating training data” with the specific example of “providing fake evaluations when rating the results of an AI-generated test report”. The question describes this attack vector: an attacker is injecting falsified test (execution) results into the training dataset, which directly manipulates the training data to compromise the LLM’s ability to accurately recommend optimal test coverage strategies. | GenAI-3.2.2 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **25** | a | Considering:<br>• Context Manipulation is based on sending requests designed to extract confidential training data. (1C)<br>• Request manipulation is based on introducing an image that disrupts the LLM output. (2D)<br>• Data poisoning is based on manipulating recommendations through fake evaluations used in training. (3A)<br>• Malicious code generation is based on manipulating a LLM to generate backdoors (e.g., external command calls) during use. (4B)<br><br>Thus:<br>a) Is correct.<br>b) Is not correct.<br>c) Is not correct.<br>d) Is not correct. | GenAI-3.2.2 | K2 | 1 |
| **26** | b | a) Is not correct. While evaluating outputs by comparing multiple LLMs is a useful practice, in this case it focuses on output quality in terms of accuracy without addressing data privacy risks.<br>b) Is correct. Anonymization is an effective strategy to mitigate data privacy risks.<br>c) Is not correct. Unrestricted access to sensitive data increases the risk of sensitive data breach.<br>d) Is not correct. Disabling encryption of sensitive data weakens security and increases the risk of sensitive data theft. | GenAI-3.2.3 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **27** | c | a) Is not correct. Image generation is by far much more energy- and $CO_2$-intensive than text generation. Energy consumption and $CO_2$ emissions are typically correlated unless a key factor (like energy source differences) is specified.<br>b) Is not correct. On the contrary, GenAI-powered searches consume considerably more energy than traditional web searches.<br>c) Is correct. Image generation requires significantly more computational resources than text generation, making it much more energy intensive.<br>d) Is not correct. The cumulative impact of text generation across millions of users is not negligible. | GenAI-3.3.1 | K2 | 1 |
| **28** | b, d | a) Is not correct. ISO/IEC 25010:2023 is a standard (not mentioned in the syllabus, but mentioned in the CTFL syllabus) that defines a product quality model that relate to quality properties of ICT/software products. It does not address the use of GenAI in software testing.<br>b) Is correct. ISO/IEC 23053:2022 is a standard (mentioned in section 3.4.1 of the syllabus) that provides a framework for data quality, transparency, and fault tolerance when using GenAI for testing.<br>c) Is not correct. ISO/IEC/IEEE 29119-2:2021 is a part (volume) of the standard (not mentioned in the syllabus, but mentioned in the CTFL syllabus) that deals with test processes. It does not address the use of GenAI in software testing.<br>d) Is correct. ISO/IEC 42001:2023 is a standard (mentioned in section 3.4.1 of the syllabus) that specifies requirements for managing AI-based systems within an organization, providing best practices for GenAI in software testing and promoting consistency and reliability.<br>e) Is not correct. ISO/IEC/IEEE 29119-3:2021 is a part (volume) of the standard (not mentioned in the syllabus, but mentioned in the CTFL syllabus) that deals with test documentation. It does not address the use of GenAI in software testing. | GenAI-3.4.1 | K1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **29** | a | a) Is correct. The back-end is responsible for retrieving data from relational and vector databases, combining it with user input, and preparing the prompt tailored for the LLM.<br>b) Is not correct. The front-end serves as the user interface for submitting input, but it does not prepare the prompt for the LLM.<br>c) Is not correct. The authentication component ensures secure access but is unrelated to data retrieval or prompt preparation.<br>d) Is not correct. The post-processing component refines the output generated by the LLM but does not handle prompt preparation. | GenAI-4.1.1 | K2 | 1 |
| **30** | a | a) Is correct. RAG automatically retrieves the relevant specifications and historical test-case chunks and feeds them to the LLM, which then generates accurate, context-aware test cases.<br>b) Is not correct. RAG is better used when asked to retrieve specific targeted information from the vector database and not the whole dataset.<br>c) Is not correct. Manually reviewing and refining the query adds unnecessary effort, which contradicts RAG's design to automate retrieval and use retrieved data dynamically in the LLM's response generation.<br>d) Is not correct. Relying on the LLM's internal data ignores RAG's core capability of enhancing responses by integrating external, up-to-date information. | GenAI-4.1.2 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **31** | c | a) Is not correct. Autonomous and semi-autonomous agents can be implemented as single-agent and/or multi-agent systems to improve both efficiency and quality in test processes. However, the key enhancements that these agents bring refer to increased efficiency and quality in test processes due to their ability to balance autonomy with human oversight.<br>b) Is not correct. While autonomous and semi-autonomous agents incorporate verifications to ensure quality, these verifications are also designed to be efficient. The use of automated checks actually aims to streamline the test processes and enhance quality without sacrificing efficiency (and vice versa) which, in turn, is also enhanced.<br>c) Is correct. Autonomous agents increase efficiency by executing test tasks within the test processes with minimal human intervention and by using automated checks designed to be efficient. Semi-autonomous agents involve strategic human oversight, which ensures the quality of test results. These two types of agents allow implementing a balanced approach to achieve both efficiency and quality by leveraging their ability to operate with varying levels of human interaction.<br>d) Is not correct. The complete elimination of verification is neither realistic nor desirable. Verification remains crucial even when using advanced technologies like autonomous and semi-autonomous LLM-powered agents. | GenAI-4.1.3 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **32** | b | a) Is not correct. This is a correct description of fine-tuning, which enhances a model's performance in a specific domain through task-specific training.<br>b) Is correct. This statement is incorrect because fine-tuning does not replace the model's general knowledge, and overfitting remains a potential challenge.<br>c) Is not correct. This is a correct description, as fine-tuning involves parameter adjustment using targeted datasets.<br>d) Is not correct. This is a correct description of the challenges associated with fine-tuning, as high-quality datasets are crucial for effective adaptation. | GenAI-4.2.1 | K2 | 1 |
| **33** | b | a) Is not correct. Large Language Model Operations (LLMOps) focuses on managing LLMs, not preventing their use.<br>b) Is correct. LLMOps manages LLMs across their lifecycle, addressing privacy, security, and cost for testing.<br>c) Is not correct. LLMOps applies to various applications, not just chatbot-based solutions.<br>d) Is not correct. LLMOps does not aim to fully automate all test tasks or eliminate human oversight. | GenAI-4.2.2 | K2 | 1 |
| **34** | d | a) Is not correct. Using unapproved GenAI tools does not enforce organizational data policies.<br>b) Is not correct. Shadow AI introduces risks related to unclear licensing agreements, which can lead to intellectual property disputes.<br>c) Is not correct. Shadow AI increases, rather than reduces, the risk of intellectual property disputes.<br>d) Is correct. Unapproved AI tools often lack robust security measures, increasing the risk of data breaches or unauthorized access. | GenAI-5.1.1 | K1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **35** | b | a) Is not correct. While certifications for specific LLMs could be useful (and more available in the future), training programs should be designed to ensure that test team members have the technical skills necessary to use GenAI tools effectively.<br>b) Is correct. Selecting LLMs that are compatible with existing test infrastructure and scalability requirements is a critical aspect of a GenAI strategy. Test tools and test environments are fundamental elements of test infrastructure.<br>c) Is not correct. Data quality and structured, secure input are crucial for achieving reliable results with GenAI. It is necessary to have an adequate amount of high-quality input data (according to the test objectives), not as much data as possible.<br>d) Is not correct. Section 5.1.2 states that a GenAI strategy for software testing must collect task-specific metrics to measure the effectiveness of LLM outputs. The metrics listed in section 2.3.1 (e.g., relevance, execution-success rate, time efficiency) are tailored to test tasks rather than borrowed wholesale from classical supervised-machine learning. | GenAI-5.1.2 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **36** | b | a) Is not correct. A key selection criterion involves evaluating the model's performance for the test tasks against the organization's benchmarks using metrics such as those presented in the syllabus itself (see the syllabus in section 5.1.3). The specified criterion may be relevant if the test tasks involve code generation. However, it is not applicable for other types of test tasks.<br>b) Is correct. A key selection criterion involves evaluating recurring costs (see the syllabus in section 5.1.3), and this answer refers to a typical example of a recurring cost.<br>c) Is not correct. Evaluating the LLM against the publicly available community benchmarks to ensure full compatibility with them is not one of the key criteria mentioned in the syllabus (in section 5.1.3) for selecting an appropriate LLM for specific test tasks.<br>d) Is not correct. A key selection criterion involves evaluating recurring costs (see the syllabus in section 5.1.3), but this answer refers to a typical example of a non-recurring cost. | GenAI-5.1.3 | K2 | 1 |

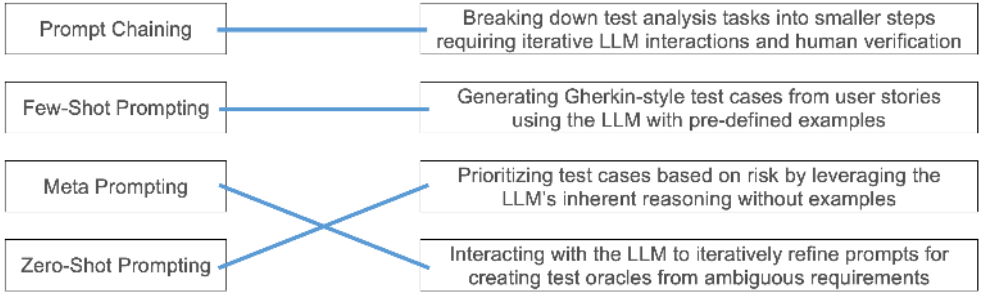
| | | | | | |
|---|---|---|---|---|---|
| **37** | a | The syllabus describes the following 3 key phases:<br>• Discovery<br>• Initiation and usage definition<br>• Utilization and iteration<br><br>Thus:<br>a) Is correct: explicitly mentions all these 3 key phases.<br>b) Is not correct: does not explicitly mention any of these 3 phases (only mentions some objectives and/or activities associated with these phases).<br>c) Is not correct: does not explicitly mention any of these 3 phases (only mentions some objectives and/or activities associated with these phases).<br>d) Is not correct: does not explicitly mention any of these 3 phases (only mentions some objectives and/or activities associated with these phases). | GenAI-5.1.4 | K1 | 1 |

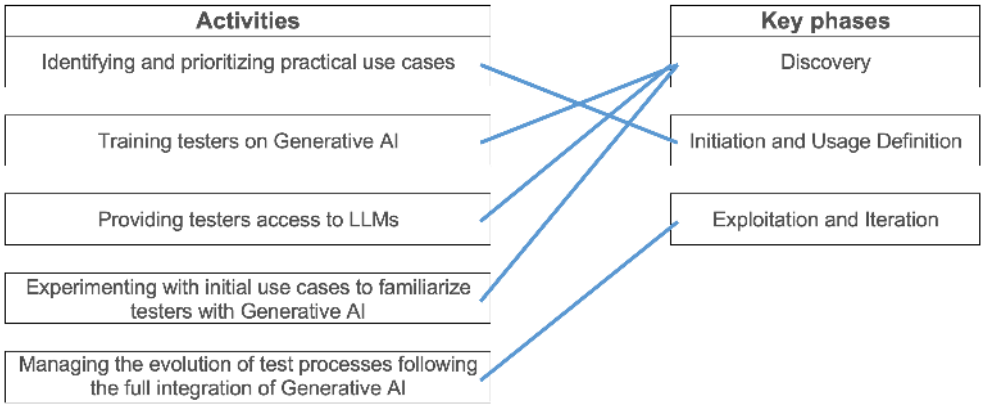
| | | | | | |
|---|---|---|---|---|---|
| **38** | c | a) Is not correct. Hallucinations in LLMs are intrinsic challenges with current AI technologies, and testers cannot prevent hallucinations and reasoning errors from occurring. Instead, testers should be able to (identify and) mitigate the risks of hallucinations and reasoning errors (and also biases) when testing with GenAI.<br>b) Is not correct. While proficiency in test automation is valuable, it does not specifically address the integration of GenAI into test processes.<br>c) Is correct. The syllabus names "assessing LLM capabilities" as a key competency. This naturally entails comparing candidate models and selecting the one that can be adapted or fine-tuned for particular test tasks. That is a direct example of knowledge/skills testers need to work effectively with LLMs.<br>d) Is not correct. It refers to skills that are required for developing the LLMs such as AI researchers and data scientists, not testers using GenAI for test tasks. These testers focus on ensuring the quality of the test data they directly use when interacting with LLMs. | GenAI-5.2.1 | K2 | 1 |
| **39** | c | a) Is not correct. While external expert courses with hands-on practice can be beneficial, mainly relying on them diminishes the importance of internal practice and community building. Moreover, integrating AI into all daily test tasks in a 'big-bang' fashion is not recommended.<br>b) Is not correct. While experimentation is part of the learning process, independent experimentation without structure may not lead to consistent or effective skill development.<br>c) Is correct. Focusing on developing practical skills through structured activities, peer learning, and knowledge-sharing communities is a recommended approach.<br>d) Is not correct. Mainly relying on theoretical courses from external experts diminishes the importance of developing practical skills and know-how through sharing within the organization. | GenAI-5.2.2 | K1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **40** | a | a) Is correct. Testers evolve into AI-assisted testing specialists, refining prompts and verifying AI outputs.<br>b) Is not correct. The responsibilities of the test managers are updated to include the development of an AI-based test strategy, AI-based risk management, and oversight of AI-based test processes. Thus, their focus is still on test management and not on understanding the (technically complex) inner workings of GenAI technologies.<br>c) Is not correct. Testers do not shift their focus to overseeing AI-based test processes. This oversight is the responsibility of test managers.<br>d) Is not correct. Test managers must balance human and AI capabilities to achieve efficient results. | GenAI-5.2.3 | K1 | 1 |

# Appendix: Answers to Additional Questions

<table>
<tr><th>Question Number (#)</th><th>Correct Answer</th><th>Explanation / Rationale</th><th>Learning Objective (LO)</th><th>K-Level</th><th>Number of Points</th></tr>
<tr><td>A1</td><td>N/A</td><td>The RAG process can be summarized as follows:<br>1. Breaking documents into chunks for focused retrieval.<br>2. Cleaning and processing chunks for consistency.<br>3. Encoding chunks into embeddings and storing them in a vector database.<br>4. Retrieving relevant chunks during user query processing.<br>5. Generating responses by combining retrieved data with the LLM’s knowledge.<br><br>Thus, the correct answer is:<br><br><table><tr><th colspan="2">Answer</th></tr><tr><td>First</td><td>Break large documents into smaller chunks</td></tr><tr><td></td><td>Clean and process document chunks</td></tr><tr><td></td><td>Store embeddings in a vector database</td></tr><tr><td></td><td>Retrieve relevant chunks based on semantic similarity</td></tr><tr><td>Last</td><td>Generate a response using retrieved chunks and the LLM</td></tr></table></td><td>GenAI-4.1.2</td><td>K2</td><td>1</td></tr>
</table>

| | | | | | |
|---|---|---|---|---|---|
| **A2** | N/A | Considering:<br>• Prompt Chaining - 1st item on the right: Directly references iterative LLM use for test analysis tasks (Section 2.2.1).<br>• Few-Shot Prompting – 2nd item on the right: Explicitly ties to LLM-generated test cases with examples (Section 2.2.2b).<br>• Meta Prompting (C) – 4th item on the right: Highlights LLM collaboration for prompt refinement (Section 2.1.2).<br>• Zero-Shot Prompting (D) – 3rd item on the right: Emphasizes LLM's intrinsic reasoning for prioritization (Section 2.1.2).<br><br>Thus, the correct answer is:<br>  <br> | GenAI-2.1.2 | K2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| **A3** | N/A | Considering:<br>• Discovery – Activities include training test teams on GenAI concepts, providing access to LLMs/SLMs, and experimenting with initial use cases to familiarize testers with GenAI and build confidence<br>• Initiation and Usage Definition – Focuses on identifying and prioritizing practical use cases for GenAI in software testing<br>• Utilization and Iteration – Continuous monitoring of the progress of GenAI for software testing and related tools is in place, as well as measurement and management of the transformation to ensure sustainable benefits and scalability<br><br>Thus, the correct answer is:<br>  <br> | GenAI-5.1.4 | K1 | 1 |