

Sample Exam – Questions

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



Copyright Notice

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

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

All rights reserved.

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

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

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

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

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

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

Document Responsibility

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

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 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 – Questions 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

Copyright Notice	2
Document Responsibility	2
Acknowledgements	2
Revision History	3
Table of Contents	4
Introduction	6
Purpose of this document	6
Instructions	6
Questions	7
Question #1 (1 Point)	7
Question #2 (1 Point)	7
Question #3 (1 Point)	7
Question #4 (1 Point)	8
Question #5 (1 Point)	8
Question #6 (1 Point)	8
Question #7 (1 Point)	9
Question #8 (1 Point)	9
Question #9 (1 Point)	9
Question #10 (1 Point)	10
Question #11 (1 Point)	10
Question #12 (2 Points)	10
Question #13 (2 Points)	11
Question #14 (2 Points)	12
Question #15 (2 Points)	13
Question #16 (2 Points)	13
Question #17 (1 Point)	14
Question #18 (1 Point)	14
Question #19 (1 Point)	15
Question #20 (2 Points)	15
Question #21 (1 Point)	15
Question #22 (1 Point)	15
Question #23 (1 Point)	16
Question #24 (1 Point)	16
Question #25 (1 Point)	16
Question #26 (1 Point)	17
Question #27 (1 Point)	17
Question #28 (1 Point)	17
Question #29 (1 Point)	17
Question #30 (1 Point)	18
Question #31 (1 Point)	18
Question #32 (1 Point)	19
Question #33 (1 Point)	19
Question #34 (1 Point)	19
Question #35 (1 Point)	19
Question #36 (1 Point)	20
Question #37 (1 Point)	20
Question #38 (1 Point)	20
Question #39 (1 Point)	21
Question #40 (1 Point)	21
Appendix: Additional Questions	22
Question #A1 (1 Point)	22
Question #A2 (1 Point)	22



Question #A3 (1 Point)..... 23

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:

- Questions¹, including for each question:
 - Any scenario needed by the question stem
 - Point value
 - Response (answer) option set
- Additional questions, including for each question [does not apply to all sample exams]:
 - Any scenario needed by the question stem
 - Point value
 - Response (answer) option set
- *Answers, including justification are contained in a separate document*

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

Questions

Question #1 (1 Point)

Match each type of AI technology (1-4) with its CORRECT description (A-D):

1. Symbolic AI
 2. Classical machine learning
 3. Deep learning
 4. Generative AI
-
- A. Uses neural networks to automatically learn features from data.
 - B. Uses rule-based systems to mimic human decision-making.
 - C. Uses deep learning to create new data by learning from its training data.
 - D. Uses a data-driven approach that requires feature selection.

- a) 1D, 2B, 3A, 4C
- b) 1D, 2C, 3B, 4A
- c) 1C, 2B, 3D, 4A
- d) 1B, 2D, 3A, 4C

Select ONE option.

Question #2 (1 Point)

Consider the realm of Large Language Models (LLMs). Which of the following options BEST explains why context window limitations affect LLM's text processing capabilities?

- a) Because context windows restrict temporal processing sequences, preventing LLMs from maintaining chronological consistency across extended text analysis.
- b) Because context windows prevent cross-referencing capabilities, limiting LLMs' ability to connect information across different document sources simultaneously.
- c) Because context windows force LLMs to discard earlier information, which may contain relevant details needed for understanding later content.
- d) Because context windows constrain parsing granularity levels, restricting LLMs from adjusting between character-level and document-level analysis approaches.

Select ONE option.

Question #3 (1 Point)

Which of the following statements BEST describes tokenization in processing text for LLMs?

- a) Tokenization converts tokens into high-dimensional vectors to capture their meaning.
- b) Tokenization creates the building blocks used to understand and generate text.
- c) Tokenization generates contextually appropriate responses using neural networks.
- d) Tokenization predicts the next token in a sequence based on learned relationships.

Select ONE option.

Question #4 (1 Point)

In the context of software testing, which of the following statements (i-v) about foundation, instruction-tuned, and reasoning LLMs are CORRECT?

- i. Foundation LLMs excel at generating test cases from high-level requirements without structured input.
 - ii. Reasoning LLMs excel at creating test scripts that strictly follow predefined organizational templates.
 - iii. Instruction-tuned LLMs excel at autonomously prioritizing test execution based on real-time user feedback.
 - iv. Reasoning LLMs excel at synthesizing data from defect reports to detect trends and prioritize test efforts.
 - v. Instruction-tuned LLMs excel at generating test cases that adhere to Gherkin language syntax.
- a) i, ii, and iii
 - b) ii, iii, and iv
 - c) i, ii, and v
 - d) iv, and v

Select ONE option.

Question #5 (1 Point)

Which of the following statements BEST describes the relation between multimodal LLMs and vision-language models?

- a) Multimodal LLMs are a subset of vision-language models designed to handle diverse inputs.
- b) Vision-language models are a subset of multimodal LLMs focusing on visual and textual data.
- c) Vision-language models are unrelated to multimodal LLMs and focus only on the user interface.
- d) Multimodal LLMs and vision-language models are interchangeable terms.

Select ONE option.

Question #6 (1 Point)

Which TWO of the following options represent key capabilities of LLMs in test tasks?

- a) Identifying ambiguities and inconsistencies in requirements.
- b) Generating complete application code for deployment.
- c) Automating the execution of all test scripts without human intervention.
- d) Performing exploratory testing on software applications.
- e) Creating diverse test data with various combinations and boundary values.

Select TWO options.

Question #7 (1 Point)

Which of the following statements BEST explains the difference between AI chatbots and LLM-powered testing applications in the context of software testing?

- a) AI chatbots are more suited for specific test tasks, while LLM-powered testing applications focus on ad hoc interactions.
- b) Both AI chatbots and LLM-powered testing applications are designed to perform identical tasks without any configuration differences.
- c) LLM-powered testing applications rely on conversational prompts, while AI chatbots require integration into test tools and test processes.
- d) AI chatbots offer conversational interfaces for ad hoc test tasks, while LLM-powered testing applications provide customized solutions for specific test tasks.

Select ONE option.

Question #8 (1 Point)

A tester is examining a structured prompt used to obtain LLM assistance for performance test analysis. One of the components of this prompt reads: “Test reports from performance testing tools, system monitoring logs during peak usage periods, and application performance benchmarks from previous releases”.

In which component of the six-part prompt structure would this description MOST LIKELY appear?

- a) Context
- b) Input data
- c) Constraints
- d) Output format

Select ONE option.

Question #9 (1 Point)

A tester wants an LLM to analyze a requirements specification for potential defects. In the structured prompt the tester is using, one line reads: “The potential defects must be provided in a markdown table with the following columns: ID, requirement reference, defect type, description, severity”.

In which component of the six-part prompt structure would this line MOST LIKELY appear?

- a) Instructions
- b) Constraints
- c) Output format
- d) Context

Select ONE option.

Question #10 (1 Point)

Which of the following BEST differentiates prompt chaining, few-shot prompting, and meta prompting techniques?

- a) Prompt chaining focuses on providing examples, few-shot prompting breaks tasks into subtasks, and meta prompting refines prompts manually.
- b) Few-shot prompting provides guidance with examples, prompt chaining breaks tasks into multiple prompts, and meta prompting allows the model to iteratively refine its own prompts.
- c) Meta prompting emphasizes breaking down tasks into steps, prompt chaining uses examples, and few-shot prompting focuses on manual optimization of prompts.
- d) Prompt chaining provides guidance without examples, few-shot provides guidance with examples, and meta prompting relies on tester-defined prompts.

Select ONE option.

Question #11 (1 Point)

What is the primary function of a system prompt in interactions with LLMs?

- a) To provide a framework for the LLM behavior for the entire conversation.
- b) To provide specific questions or instructions from the user to the LLM.
- c) To adjust dynamically with each user interaction and set the conversation's context.
- d) To include visible input from the user and set rules for the conversation.

Select ONE option.

Question #12 (2 Points)

You are tasked with applying the following test approach to a set of stable requirements for a new project: generate test conditions, prioritize them based on risk level, and identify potential coverage gaps. The requirements have already been thoroughly reviewed for defects.

Which of the following sequences of steps (i-v) should you follow to effectively apply Generative AI to implement this test approach using a prompt chaining technique to implement this test approach?

- i. Submit the requirements to the LLM and prompt it to produce test conditions based on those requirements.
 - ii. Provide the test conditions to the LLM, ensuring it understands the context for prioritization, and prompt it to prioritize those test conditions accordingly.
 - iii. Provide the prioritized test conditions to the LLM and prompt it to analyze them to determine whether all aspects of the requirements are addressed in the test conditions.
 - iv. Submit the requirements to the LLM and prompt it to produce prioritized test conditions that address all aspects of the requirements.
 - v. Submit the requirements to the LLM and prompt it to detect inconsistencies and ambiguities in those requirements.
-
- a) i, ii, and iii
 - b) iv, and ii
 - c) i, iii, and v
 - d) v, and iv

Select ONE option.

Question #13 (2 Points)

Consider applying the few-shot structured prompting technique to generate Gherkin-style test cases (i.e., scenario-based) for the following user story and acceptance criterion:

- User story: "As a user, I want to reset my password so that I can regain access to my account if I forget it."
- Acceptance criterion: "When a user submits a registered email address then they receive a password reset email."

You can rely on predefined examples that include user stories, acceptance criteria, and Gherkin-style test cases. Your task is to create a prompt to guide the LLM in generating accurate test cases aligned with the acceptance criterion for the user story above.

Which one of the following prompts is BEST suited to this task?

a) Prompt A

Role: Act as a test analyst.

Context: You are testing a password reset functionality.

Instruction: Generate Gherkin-style test cases for the user story and acceptance criterion. using the following predefined examples as a guide: << predefined examples >>.

Input Data: <<< user story >>> and <<< acceptance criterion >>>.

Constraints: Rely on best practices to create test cases.

Output Format: Generate test cases with expected results.

b) Prompt B

Role: Act as a test analyst specializing in Gherkin-style test cases.

Context: You are testing a password reset functionality.

Instruction: Generate Gherkin-style test cases for the user story and acceptance criterion, using the following predefined examples as a guide: << predefined examples >>.

Input Data: <<< user story >>> and <<< acceptance criterion >>>.

Constraints: Use "Given-When-Then" syntax and ensure alignment with the acceptance criterion.

Output Format: Respect the given Gherkin-style test case format.

c) Prompt C

Role: Act as a test analyst.

Context: You are testing a password reset functionality.

Instruction: Generate Gherkin-style test cases for the user story and acceptance criterion. Rely on best practices to create test cases.

Input Data: <<< user story >>> and <<< acceptance criterion >>>.

Constraints: Use "Given-When-Then" syntax and ensure alignment with the acceptance criterion.

Output Format: Respect the given Gherkin-style test case format.

d) Prompt D

Role: Act as a test analyst.

Context: You are testing a password reset functionality.

Instruction: Generate at least two Gherkin-style test cases for the user story and acceptance criterion. Focus on edge cases.

Input Data: <<< user story >>> and <<< acceptance criterion >>>.

Constraints: Ensure all test cases follow "Given-When-Then" syntax.

Output Format: Respect the given Gherkin-style test case format.

Select ONE option.

Question #14 (2 Points)

You are tasked with applying structured prompting to analyze regression test results. Here is an initial draft of the prompt:

Role: Act as a test analyst.

Context: Analyze raw regression test results from a recent test execution cycle.

Instruction: Identify discrepancies in the test results.

Input Data: Use the attached file containing raw test results.

Constraints: Use the known anomalies list for cross-checking.

Output Format: Provide a list of discrepancies using a table format.

You are asked to improve this prompt. Which of the following improvements would BEST align the prompt with structured prompt engineering best practices for comprehensive regression test report analysis?

- a) Add a step to cluster similar issues and cross-check findings against the known anomalies list.
- b) Specify that the role is a regression test analyst specializing in actionable insights.
- c) Expand the instruction to include separating expected results and actual results, clustering issues, and highlighting discrepancies.
- d) Include references to regression testing principles such as "Given-When-Then" in the constraints.

Select ONE option.

Question #15 (2 Points)

You are using an LLM to assist in preparing actionable test metrics from raw data. The metrics include test progress, defect trends, and coverage, which are graphically displayed and explained with text. Your goal is to improve the test process to ensure the generated metrics are accurate, actionable, and easily interpretable by stakeholders.

Here is an initial draft of a prompt used to instruct the AI:

Role: Act as a test manager.

Context: You are provided with raw data from test tools.

Instruction: Generate test progress metrics, defect trend metrics, and coverage metrics from the raw data.

Input Data: Use the attached file containing raw test results.

Constraints: Ensure that the output is concise and understandable.

Output Format: Display metrics on a dashboard.

You are asked to improve this prompt. Which of the following improvements would BEST enhance the LLM's ability to produce accurate and actionable metrics?

- a) Specify that the role is a test manager focusing on actionable insights and decision support, ensuring comprehensive analysis of test data.
- b) Add an instruction to include potential risks identified from the trends in the generated metrics, along with their impact assessment and priority levels.
- c) Expand the output format to include a plain-language summary that interprets the metrics and outlines next steps for stakeholders.
- d) Emphasize constraints that the output is also easily interpretable by stakeholders, using clear language and avoiding technical jargon throughout the response.

Select ONE option.

Question #16 (2 Points)

Your goal is to create test cases for an AI-based system that suffers from the test oracle problem, preventing you from determining expected results. You can only count on a few existing test cases with known expected results. Through appropriate analysis, you have identified a set of well-defined transformation rules that specify how changes to inputs affect expected results. These rules can be applied to all existing test cases. You have decided to rely on Generative AI, providing a given LLM with the following information: the existing test cases with their inputs and expected results, a clear description of the transformation rules, and guidelines for generating additional test cases by precisely applying these rules to the relevant existing test cases. With the specified information, the chosen LLM can directly generate additional test cases in line with your expectations.

Which of the following prompting techniques is BEST suited to achieve your goal in this scenario?

- a) Few-shot prompting
- b) Prompt chaining
- c) Meta prompting
- d) Zero-shot prompting

Select ONE option.

Question #17 (1 Point)

You are leveraging Generative AI to assist in testing an entertainment software application. The Generative AI model generates test cases for user interaction scenarios, test scripts for API interactions, and synthetic test data to address edge cases.

To effectively evaluate the Generative AI model's performance and to refine prompts, which combination of metrics and actions BEST ensures comprehensive assessment and improvement?

- a) Evaluate the diversity of test cases to ensure varied input scenarios and use test execution success rate to validate the functionality of generated API test scripts.
- b) Apply accuracy and completeness metrics to validate test cases against entertainment software requirements and rely on time efficiency to compare AI-generated test scripts with manual test efforts.
- c) Focus on precision to ensure generated test data meets entertainment software compliance standards, while contextual fit and test execution success rate assesses the alignment and usability of test scripts.
- d) Prioritize relevance and contextual fit for all outputs to maintain consistency with entertainment software requirements and include diversity metrics to expand edge case coverage.

Select ONE option.

Question #18 (1 Point)

Which of the following techniques for evaluating and iteratively refining prompts is BEST suited for determining why an LLM consistently generates test cases with wrong expected results that contradict the input requirements, thereby providing insights to optimize the prompt and prevent similar errors?

- a) Output analysis
- b) A/B testing of prompts
- c) Adjusting prompt length and specificity
- d) Integrating user feedback

Select ONE option.

Question #19 (1 Point)

What is a hallucination in the context of LLM outputs?

- a) A logical error where the LLM fails to follow a multi-step reasoning process accurately.
- b) A bias in the LLM output caused by the training data favoring certain perspectives.
- c) A generation of irrelevant or factually incorrect output by the LLM for a given task.
- d) A limitation of the LLM to understand non-English perspectives in test generation tasks.

Select ONE option.

Question #20 (2 Points)

You are using Generative AI to create test cases for an e-commerce (e-shop) application. The following features have been explicitly mentioned in the project briefing:

- cart management
- discount code application
- order confirmation email generation

Based on these details, which of the following AI-generated test cases MOST LIKELY represents a hallucination?

- a) Verify that a user can add multiple items to their cart and proceed to checkout.
- b) Verify that a user cannot apply an expired discount code during checkout.
- c) Verify that a user receives a confirmation email after successfully placing an order.
- d) Verify that a user can create a wishlist to save favorite items for later.

Select ONE option.

Question #21 (1 Point)

Which of the following options refers to a benefit that is MOST directly associated with using clear and structured input data formats when working with LLMs for test tasks?

- a) Helps reduce the effort to fine-tune the LLMs for test tasks.
- b) Helps LLMs generate less ambiguous outputs for test tasks.
- c) Helps LLMs generate more context-relevant outputs for test tasks.
- d) Helps LLMs generate more creative outputs for test tasks.

Select ONE option.

Question #22 (1 Point)

Which strategy can help reduce variability in LLM outputs by narrowing the probability distribution during inference?

- a) Increasing the learning rate.
- b) Lowering the temperature setting.
- c) Increasing the random seed.
- d) Lowering the random seed.

Select ONE option.

Question #23 (1 Point)

Which of the following statements about data privacy concerns related to using Generative AI for software testing is INCORRECT?

- a) Generative AI can unintentionally expose sensitive data through its outputs.
- b) Generative AI tools may store and process sensitive data without explicit user consent, leading to misuse.
- c) Using Generative AI tools without adhering to data protection regulations, such as General Data Protection Regulation (GDPR), can lead to legal disputes.
- d) An LLM is likely to expose real sensitive data if it hallucinates while generating synthetic test data, regardless of the data it was trained on.

Select ONE option.

Question #24 (1 Point)

An attacker injects falsified test results into the training dataset of an LLM intended to recommend optimal test coverage strategies. What type of attack vector does this description BEST refer to?

- a) Malicious code generation
- b) Context Manipulation
- c) Request manipulation
- d) Data poisoning

Select ONE option.

Question #25 (1 Point)

Match each type of attack vector against an LLM (1-4) with the corresponding example (A-D):

- 1. Context Manipulation
 - 2. Request manipulation
 - 3. Data poisoning
 - 4. Malicious code generation
-
- A. An attacker maliciously modifies the data associated with traceability links between requirements and test cases into the dataset used for fine-tuning an LLM, compromising its accuracy in generating test cases from requirements.
 - B. An attacker maliciously crafts and provides deceptive prompts that induce an LLM, fine-tuned to assist testers in automated test script generation, to produce vulnerable test scripts with hidden security flaws.
 - C. An attacker maliciously provides large specially crafted prompts that induce an LLM, fine-tuned to assist testers in generating test cases, to accidentally reveal confidential API keys inherited from past test projects.
 - D. An attacker maliciously submits carefully modified reference screenshots into a visual testing framework that uses an LLM for comparative visual analysis, to trick the LLM into systematically ignoring genuine UI issues during regression testing.

- a) 1C, 2D, 3A, 4B
- b) 1B, 2D, 3A, 4C
- c) 1D, 2C, 3B, 4A
- d) 1C, 2B, 3D, 4A

Select ONE option.

Question #26 (1 Point)

Which of the following strategies BEST addresses data privacy risks in the context of Generative AI-powered software testing?

- a) Using multiple LLMs to evaluate and compare test results for improved accuracy.
- b) Replacing sensitive test data with an anonymized version of the same.
- c) Allowing unrestricted access to sensitive test data to improve Generative AI model training.
- d) Disabling encryption of sensitive test data to streamline data storage and transmission processes.

Select ONE option.

Question #27 (1 Point)

Which of the following options about the impact of LLM usage on energy consumption and CO₂ emission is CORRECT?

- a) Image generation tasks consume substantially more energy than text generation tasks, but produce fewer CO₂ emissions.
- b) Generative AI-powered searches consume significantly less energy than traditional web searches due to their optimized algorithms.
- c) Image generation tasks consume substantially more energy than text generation tasks due to their higher computational complexity.
- d) Text generation tasks consume very little energy, allowing them to be performed by millions of users without significant energy consumption.

Select ONE option.

Question #28 (1 Point)

Which TWO of the following standards, or parts of them, are MOST relevant to the use of Generative AI in software testing?

- a) ISO/IEC 25010:2023
- b) ISO/IEC 23053:2022
- c) ISO/IEC/IEEE 29119-2:2021
- d) ISO/IEC 42001:2023
- e) ISO/IEC/IEEE 29119-3:2021

Select TWO options.

Question #29 (1 Point)

Which of the following components of an LLM-powered testing application is responsible for combining user input with structured and semantically similar data to prepare a prompt for the LLM?

- a) Back-end
- b) Front-end
- c) Authentication component
- d) Post-processing component

Select ONE option.

Question #30 (1 Point)

You are a tester working on a banking application that includes features such as user login, account management, and secure transactions. The system documentation, including API specifications and security requirements, is stored in a vector database, while historical test cases are stored in a relational database. Your task is to generate test cases using a Retrieval-Augmented Generation (RAG) framework to ensure alignment with the latest specifications and requirements.

Which of the following options represents the MOST appropriate use of the RAG framework in this scenario?

- a) Submit a query specifying one function to be tested. The RAG framework will retrieve relevant specifications and requirements from the vector database, combine them with historical test cases, and automatically generate accurate and context-aware test cases through the LLM.
- b) Submit a query specifying all functions to be tested. The RAG framework will retrieve relevant specifications and requirements from the vector database, combine them with historical test cases, and automatically generate accurate and context-aware test cases through the LLM.
- c) Use the RAG framework to retrieve historical test cases from the relational database and security requirements from the vector database. Manually review the retrieved information before refining the query for the LLM to generate targeted test cases.
- d) Rely on the LLM's internal training data to generate test cases while using the RAG framework or reference retrieval without directly integrating retrieved information into the generation process.

Select ONE option.

Question #31 (1 Point)

Which of the following options BEST describes the enhancements that autonomous and semi-autonomous LLM-powered agents bring to automating test processes?

- a) They can enhance both efficiency and quality in automating test processes through a balanced use of single-agent and multi-agent systems.
- b) They can enhance quality in automating test processes by adding complex verification checks although at the expense of efficiency.
- c) They can enhance both efficiency and quality in automating test processes by leveraging their ability to operate with varying levels of human interaction.
- d) They can enhance both efficiency and quality in automating test processes while eliminating the need for verification within these test processes.

Select ONE option.

Question #32 (1 Point)

Which of the following statements about fine-tuning language models for specific test tasks is INCORRECT?

- a) Fine-tuning involves training a pre-trained model on task-specific data to enhance its performance and domain knowledge.
- b) Fine-tuning equips a language model with new capabilities by replacing its general knowledge with task-specific reasoning, ensuring the absence of overfitting.
- c) Fine-tuning modifies a pre-trained model's parameters using a targeted dataset to adapt it for a specific domain or task.
- d) Fine-tuning requires high-quality, task-specific datasets to avoid biased or inaccurate results.

Select ONE option.

Question #33 (1 Point)

Which of the following BEST describes the primary focus of Large Language Model Operations (LLMOps) when deploying and managing LLMs for test tasks?

- a) Preventing reliance on Generative AI in test processes.
- b) Managing LLMs effectively across their lifecycle, including privacy, security, and cost considerations.
- c) Limiting LLM usage to chatbot-based testing solutions to reduce complexity.
- d) Fully automating all test tasks without requiring human oversight.

Select ONE option.

Question #34 (1 Point)

Which of the following statements about shadow AI is CORRECT?

- a) Shadow AI enforces compliance with organizational data policies and general AI regulations.
- b) Shadow AI eliminates the need for clear licensing agreements in AI tools.
- c) Shadow AI reduces the risk of intellectual property disputes.
- d) Shadow AI may lead to unauthorized accesses to sensitive information.

Select ONE option.

Question #35 (1 Point)

What is a key aspect to consider when defining a Generative AI strategy for software testing?

- a) Prepare training programs designed to ensure that team members obtain certifications specific to each LLM they use.
- b) Select LLMs that can be integrated appropriately with existing test environments and test tools.
- c) Ensure that as much input data as possible is available to increase the likelihood of obtaining effective LLM outputs.
- d) Collect standard supervised machine learning metrics to evaluate the effectiveness of LLM outputs.

Select ONE option.

Question #36 (1 Point)

Which of the following statements BEST describes one of the key criteria to select an appropriate LLM for specific test tasks within a test organization?

- a) A key selection criterion involves evaluating the LLM's performance for the test tasks against the publicly available benchmarks for LLMs in code generation.
- b) A key selection criterion involves evaluating recurring costs such as those associated with the computational resources required to run the LLM.
- c) A key selection criterion involves evaluating the LLM against the publicly available community benchmarks to ensure full compatibility with them.
- d) A key selection criterion involves evaluating recurring costs such as those associated with a proof of concept aimed at demonstrating the suitability of an LLM for the test tasks.

Select ONE option.

Question #37 (1 Point)

What are the key phases in the adoption of Generative AI in a test organization?

- a) Discovery, initiation and usage definition, utilization and iteration
- b) Awareness, usage prioritization, performance monitoring
- c) Planning, experimentation, evaluation and refinement
- d) Training, testing, implementation, and scaling

Select ONE option.

Question #38 (1 Point)

Which of the following options BEST refers to an example of knowledge and/or skills required for testers to work effectively with LLMs in test processes?

- a) Mastering techniques specifically aimed at preventing LLMs from hallucinating and incurring reasoning errors when performing specific test tasks.
- b) Selecting and implementing suitable test automation approaches, such as keyword-driven test automation, for automating test processes.
- c) Choosing the most appropriate LLM based on criteria such as its capability to be adapted or customized to perform specific test tasks.
- d) Ensuring the validation and test data used in the development of the LLMS are of the highest quality.

Select ONE option.

Question #39 (1 Point)

What is the BEST approach for cultivating skills within test teams to specifically support the adoption of Generative AI?

- a) Rely mainly on external expert courses with hands-on practice, aiming to integrate AI into all daily test tasks at once.
- b) Encourage independent experimentation with various LLMs without following a structured process.
- c) Adopt a hands-on, gradual learning process supported by guided exercises, peer learning, and knowledge-sharing communities.
- d) Rely mainly on theoretical courses from external experts, aiming to gradually integrate AI into daily test tasks in line with actual learning.

Select ONE option.

Question #40 (1 Point)

Which of the following answers BEST describes how the roles and responsibilities of testers and test managers within a test organization are impacted by the adoption of Generative AI for software testing?

- a) Testers shift their focus from manually designing test cases to guiding and verifying AI-generated testware.
- b) Test managers shift their focus from managing test projects to understanding the inner working of Generative AI technologies.
- c) Testers shift their focus from manually designing test cases to overseeing AI-based test processes.
- d) Test managers shift their focus from relying on people to relying solely on Generative AI to boost productivity in test tasks.

Select ONE option.

Appendix: Additional Questions

Question #A1 (1 Point)

Consider the following steps (on the left) involved in implementing RAG (Retrieval-Augmented Generation):

Steps	Answer
Clean and process document chunks	First
Generate a response using retrieved chunks and the LLM	
Break large documents into smaller chunks	
Store embeddings in a vector database	
Retrieve relevant chunks based on semantic similarity	Last

Order these steps in sequence from the first to the last.

Question #A2 (1 Point)

Match each prompting technique (on the left) to an appropriate example of its application in software testing (on the right):

Prompt Chaining	Breaking down test analysis tasks into smaller steps requiring iterative LLM interactions and human verification
Few-Shot Prompting	Generating Gherkin-style test cases from user stories using the LLM with pre-defined examples
Meta Prompting	Prioritizing test cases based on risk by leveraging the LLM's inherent reasoning without examples
Zero-Shot Prompting	Interacting with the LLM to iteratively refine prompts for creating test oracles from ambiguous requirements

Assign each individual prompting technique with an appropriate example of its application in software testing. No item may be left unpaired or paired with more than one other item.

Question #A3 (1 Point)

Consider the following activities (on the left) and key phases (on the right) involved in the adoption of Generative AI within a test organization:

Activities	Key phases
Identifying and prioritizing practical use cases	Discovery
Training testers on Generative AI	Initiation and Usage Definition
Providing testers access to LLMs	Utilization and Iteration
Experimenting with initial use cases to familiarize testers with Generative AI	
Managing the evolution of test processes following the full integration of Generative AI	

Assign each activity (item) to a key phase (group). No group can be left empty and no item can be assigned to more than one group.

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



Copyright Notice

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

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

All rights reserved.

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

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

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

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

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

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

Document Responsibility

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

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

Copyright Notice	2
Document Responsibility	2
Acknowledgements.....	2
Revision History	3
Table of Contents.....	4
Introduction	6
Purpose of this document.....	6
Instructions	6
Answer Key.....	7
Answers	8
1	8
2	9
3	9
4	10
5	11
6	11
7	12
8	12
9	13
10	13
11	14
12	15
13	16
14	16
15	17
16	17
17	18
18	18
19	19
20	19
21	20
22	20
23	21
24	22
25	23
26	23
27	24
28	24
29	25
30	25
31	26
32	27
33	27
34	27
35	28
36	29
37	30
38	31
39	31
40	32
Appendix: Answers to Additional Questions	33
A1	33
A2	34

A3..... 35

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: <ul style="list-style-type: none"> • Symbolic AI uses a rule-based system to mimic human decision-making, representing knowledge using symbols and logical rules (1B). • Classical machine learning uses a data-driven approach that requires data preparation, feature selection and model training (2D). • Deep learning uses neural networks (machine learning structures) to automatically learn features from data (3A). • Generative AI uses deep learning techniques to create new data by learning and mimicking patterns from its training data (4C). Thus: <ol style="list-style-type: none"> a) Is not correct. b) Is not correct. c) Is not correct. d) Is correct. 	GenAI-1.1.1	K1	1

2	c	<p>a) Is not correct because context windows do not control temporal sequencing but limit the scope of text that can be simultaneously considered.</p> <p>b) Is not correct because context windows affect scope within the current input, not cross-document referencing capabilities.</p> <p>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.</p> <p>d) Is not correct because context windows do not determine parsing approaches but limit the amount of text that can be processed simultaneously.</p>	GenAI-1.1.2	K2	1
3	b	<p>a) Is not correct. This describes embeddings, not tokenization.</p> <p>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”).</p> <p>c) Is not correct. This describes the general function of LLMs, not tokenization.</p> <p>d) Is not correct. This refers to how LLMs generate text, not tokenization.</p>	GenAI-1.1.2	K2	1

4	d	<p>Considering:</p> <ul style="list-style-type: none"> 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. 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. 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. 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. 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. <p>Thus:</p> <ul style="list-style-type: none"> a) Is not correct. b) Is not correct. c) Is not correct. d) Is correct. 	GenAI-1.1.3	K2	1
---	---	---	-------------	----	---

5	b	<p>a) Is not correct. Vision-language models are a subset of multimodal LLMs, not the reverse.</p> <p>b) Is correct. Vision-language models specifically integrate visual and textual data, making them a subset of multimodal LLMs.</p> <p>c) Is not correct. Vision-language models are closely related to multimodal LLMs and focus on both visual and textual data.</p> <p>d) Is not correct. Multimodal LLMs and vision-language models have distinct scopes and are not interchangeable.</p>	GenAI-1.1.4	K2	1
6	a, e	<p>a) Is correct. LLMs can analyze and clarify requirements by identifying ambiguities and inconsistencies.</p> <p>b) Is not correct. Generating complete application code is not a key capability of LLMs in test tasks.</p> <p>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.</p> <p>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.</p> <p>e) Is correct. LLMs can generate diverse test data, including combinations and boundaries.</p>	GenAI-1.2.1	K2	1

7	d	<p>a) Is not correct. AI chatbots are best suited for ad hoc interactions, not specific test tasks.</p> <p>b) Is not correct. AI chatbots and LLM-powered testing applications have distinct purposes and are not identical in functionality or configurability.</p> <p>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.</p> <p>d) Is correct. AI chatbots provide conversational interfaces for ad hoc tasks, while LLM-powered testing applications deliver customized solutions for specific needs.</p>	GenAI-1.2.2	K2	1
8	b	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	GenAI-2.1.1	K2	1

9	c	<p>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.</p> <p>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.</p> <p>c) Is correct. The given line matches the syllabus definition of output format specifications that guide how the LLM should format the output.</p> <p>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.</p>	GenAI-2.1.1	K2	1
10	b	<p>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”.</p> <p>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.</p> <p>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”.</p> <p>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.</p>	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. b) Is not correct. This describes the function of a user prompt, not a system prompt. c) Is not correct. The system prompt does not adjust dynamically; it remains constant. 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	<p>Considering:</p> <ul style="list-style-type: none"> 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. 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. 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. 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. 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). <p>Thus:</p> <ul style="list-style-type: none"> a) Is correct. b) Is not correct. c) Is not correct. d) Is not correct. 	GenAI-2.2.1	K3	2
----	---	---	-------------	----	---

13	b	<p>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.</p> <p>b) Is correct. Comprehensive and leverages predefined examples to guide the LLM.</p> <p>c) Is not correct. Lacks emphasis on using predefined examples and indicates reliance on vague best practices in defining the instructions.</p> <p>d) Is not correct. Focuses on edge cases but neglects comprehensive coverage and the use of examples for guidance.</p>	GenAI-2.2.2	K3	2
14	c	<p>a) Is not correct. Addresses clustering and cross-checking but misses other critical steps like separating test results.</p> <p>b) Is not correct. Improves role clarity but does not expand instructions or address structured steps.</p> <p>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.</p> <p>d) Is not correct. Introduces irrelevant constraints, misaligning the prompt with the task requirements.</p>	GenAI-2.2.3	K3	2

15	c	<p>a) Is not correct. Clarifies the role but does not add any concrete instruction that improves accuracy, actionability, or interpretability of the delivered metrics.</p> <p>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.</p> <p>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”).</p> <p>d) Is not correct. Merely re-states an existing constraint and gives no specific guidance on how the LLM should achieve stakeholder-level interpretability.</p>	GenAI-2.2.4	K3	2
16	a	<p>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.</p> <p>b) Is not correct. While prompt chaining could be used, it could unnecessarily complicate this very straightforward task.</p> <p>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 .</p> <p>d) Is not correct. Zero-shot prompting would not be effective since it would not leverage existing test cases as examples.</p>	GenAI-2.2.5	K3	2

17	a	<p>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).</p> <p>b) Is not correct. While accuracy and completeness are important, relying primarily on time efficiency does not fully evaluate coverage or test execution reliability.</p> <p>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.</p> <p>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.</p>	GenAI-2.3.1	K2	1
18	a	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	GenAI-2.3.2	K2	1

19	c	<p>a) Is not correct. This describes reasoning errors, not hallucinations.</p> <p>b) Is not correct. This described biases in AI output, not hallucinations.</p> <p>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.</p> <p>d) Is not correct. This describes biases due to underrepresentation in training data, not hallucinations.</p>	GenAI-3.1.1	K1	1
20	d	<p>a) Is not correct. Cart management is explicitly mentioned in the project briefing, making this test case relevant.</p> <p>b) Is not correct. Discount code application is explicitly mentioned in the project briefing, making this test case relevant.</p> <p>c) Is not correct. Order confirmation emails are explicitly mentioned in the project briefing, so this test case is valid.</p> <p>d) Is correct. Wishlist management is not mentioned in the project briefing, making this the most likely hallucinated test case.</p>	GenAI-3.1.2	K3	2

<p>21</p>	<p>b</p>	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	<p>GenAI-3.1.3</p>	<p>K2</p>	<p>1</p>
<p>22</p>	<p>b</p>	<p>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.</p> <p>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.</p> <p>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.</p>	<p>GenAI-3.1.4</p>	<p>K1</p>	<p>1</p>

23	d	<p>a) Is not correct. This is a privacy concern, as “Unintentional data exposure” relates to generative AI (GenAI) outputs revealing sensitive information accidentally.</p> <p>b) Is not correct. This is a privacy concern, as it involves the lack of control over how sensitive data is stored or processed.</p> <p>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.</p> <p>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.</p>	GenAI-3.2.1	K2	1
----	---	---	-------------	----	---

24	d	<p>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.</p> <p>b) Is not correct. Context Manipulation involves “sending requests designed to extract confidential training data”, not injecting false data into training datasets.</p> <p>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.</p> <p>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.</p>	GenAI-3.2.2	K2	1
----	---	--	-------------	----	---

<p>25</p>	<p>a</p>	<p>Considering:</p> <ul style="list-style-type: none"> • Context Manipulation is based on sending requests designed to extract confidential training data. (1C) • Request manipulation is based on introducing an image that disrupts the LLM output. (2D) • Data poisoning is based on manipulating recommendations through fake evaluations used in training. (3A) • Malicious code generation is based on manipulating a LLM to generate backdoors (e.g., external command calls) during use. (4B) <p>Thus:</p> <p>a) Is correct. b) Is not correct. c) Is not correct. d) Is not correct.</p>	<p>GenAI-3.2.2</p>	<p>K2</p>	<p>1</p>
<p>26</p>	<p>b</p>	<p>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. b) Is correct. Anonymization is an effective strategy to mitigate data privacy risks. c) Is not correct. Unrestricted access to sensitive data increases the risk of sensitive data breach. d) Is not correct. Disabling encryption of sensitive data weakens security and increases the risk of sensitive data theft.</p>	<p>GenAI-3.2.3</p>	<p>K2</p>	<p>1</p>

27	c	<p>a) Is not correct. Image generation is by far much more energy- and CO₂-intensive than text generation. Energy consumption and CO₂ emissions are typically correlated unless a key factor (like energy source differences) is specified.</p> <p>b) Is not correct. On the contrary, GenAI-powered searches consume considerably more energy than traditional web searches.</p> <p>c) Is correct. Image generation requires significantly more computational resources than text generation, making it much more energy intensive.</p> <p>d) Is not correct. The cumulative impact of text generation across millions of users is not negligible.</p>	GenAI-3.3.1	K2	1
28	b, d	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	GenAI-3.4.1	K1	1

29	a	<p>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.</p> <p>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.</p> <p>c) Is not correct. The authentication component ensures secure access but is unrelated to data retrieval or prompt preparation.</p> <p>d) Is not correct. The post-processing component refines the output generated by the LLM but does not handle prompt preparation.</p>	GenAI-4.1.1	K2	1
30	a	<p>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.</p> <p>b) Is not correct. RAG is better used when asked to retrieve specific targeted information from the vector database and not the whole dataset.</p> <p>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.</p> <p>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.</p>	GenAI-4.1.2	K2	1

31	c	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	GenAI-4.1.3	K2	1
----	---	---	-------------	----	---

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

35	b	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	GenAI-5.1.2	K2	1
----	---	--	-------------	----	---

36	b	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	GenAI-5.1.3	K2	1
----	---	--	-------------	----	---

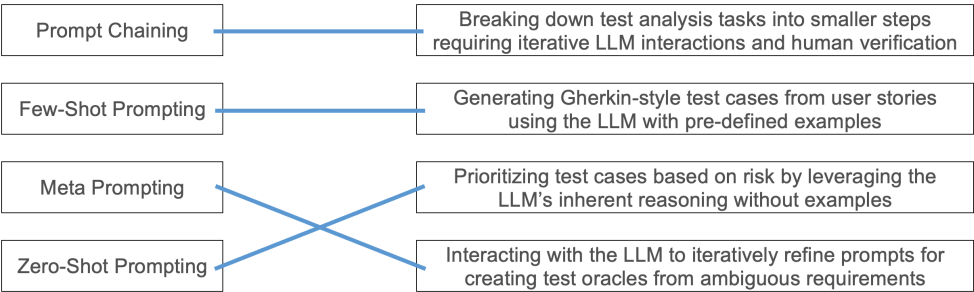
37	a	<p>The syllabus describes the following 3 key phases:</p> <ul style="list-style-type: none"> • Discovery • Initiation and usage definition • Utilization and iteration <p>Thus:</p> <p>a) Is correct: explicitly mentions all these 3 key phases.</p> <p>b) Is not correct: does not explicitly mention any of these 3 phases (only mentions some objectives and/or activities associated with these phases).</p> <p>c) Is not correct: does not explicitly mention any of these 3 phases (only mentions some objectives and/or activities associated with these phases).</p> <p>d) Is not correct: does not explicitly mention any of these 3 phases (only mentions some objectives and/or activities associated with these phases).</p>	GenAI-5.1.4	K1	1
----	---	--	-------------	----	---

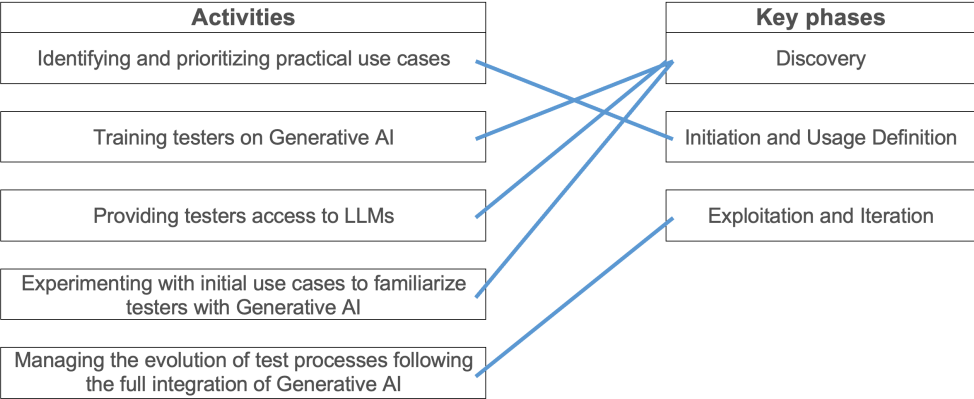
38	c	<p>a) Is not correct. Hallucinations in LLMs are intrinsic challenges with current AI technologies, and testers cannot prevent hallucinations and reasoning errors from occurring. Instead, testers should be able to (identify and) mitigate the risks of hallucinations and reasoning errors (and also biases) when testing with GenAI.</p> <p>b) Is not correct. While proficiency in test automation is valuable, it does not specifically address the integration of GenAI into test processes.</p> <p>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.</p> <p>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.</p>	GenAI-5.2.1	K2	1
39	c	<p>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.</p> <p>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.</p> <p>c) Is correct. Focusing on developing practical skills through structured activities, peer learning, and knowledge-sharing communities is a recommended approach.</p> <p>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.</p>	GenAI-5.2.2	K1	1

40	a	<p>a) Is correct. Testers evolve into AI-assisted testing specialists, refining prompts and verifying AI outputs.</p> <p>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.</p> <p>c) Is not correct. Testers do not shift their focus to overseeing AI-based test processes. This oversight is the responsibility of test managers.</p> <p>d) Is not correct. Test managers must balance human and AI capabilities to achieve efficient results.</p>	GenAI-5.2.3	K1	1
----	---	---	-------------	----	---

Appendix: Answers to Additional Questions

Question Number (#)	Correct Answer	Explanation / Rationale	Learning Objective (LO)	K-Level	Number of Points												
A1	N/A	<p>The RAG process can be summarized as follows:</p> <ol style="list-style-type: none"> 1. Breaking documents into chunks for focused retrieval. 2. Cleaning and processing chunks for consistency. 3. Encoding chunks into embeddings and storing them in a vector database. 4. Retrieving relevant chunks during user query processing. 5. Generating responses by combining retrieved data with the LLM's knowledge. <p>Thus, the correct answer is:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th colspan="2" style="text-align: center;">Answer</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">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 style="text-align: center;">Last</td> <td>Generate a response using retrieved chunks and the LLM</td> </tr> </tbody> </table>	Answer		First	Break large documents into smaller chunks		Clean and process document chunks		Store embeddings in a vector database		Retrieve relevant chunks based on semantic similarity	Last	Generate a response using retrieved chunks and the LLM	GenAI-4.1.2	K2	1
Answer																	
First	Break large documents into smaller chunks																
	Clean and process document chunks																
	Store embeddings in a vector database																
	Retrieve relevant chunks based on semantic similarity																
Last	Generate a response using retrieved chunks and the LLM																

<p>A2</p>	<p>N/A</p>	<p>Considering:</p> <ul style="list-style-type: none"> • Prompt Chaining - 1st item on the right: Directly references iterative LLM use for test analysis tasks (Section 2.2.1). • Few-Shot Prompting – 2nd item on the right: Explicitly ties to LLM-generated test cases with examples (Section 2.2.2b). • Meta Prompting (C) – 4th item on the right: Highlights LLM collaboration for prompt refinement (Section 2.1.2). • Zero-Shot Prompting (D) – 3rd item on the right: Emphasizes LLM's intrinsic reasoning for prioritization (Section 2.1.2). <p>Thus, the correct answer is:</p> 	<p>GenAI-2.1.2</p>	<p>K2</p>	<p>1</p>
------------------	------------	---	--------------------	-----------	----------

<p>A3</p>	<p>N/A</p>	<p>Considering:</p> <ul style="list-style-type: none"> • 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 • Initiation and Usage Definition – Focuses on identifying and prioritizing practical use cases for GenAI in software testing • 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 <p>Thus, the correct answer is:</p> <div style="display: flex; justify-content: space-around;"> <div data-bbox="533 735 1016 1134"> <table border="1"> <thead> <tr> <th>Activities</th> </tr> </thead> <tbody> <tr> <td>Identifying and prioritizing practical use cases</td> </tr> <tr> <td>Training testers on Generative AI</td> </tr> <tr> <td>Providing testers access to LLMs</td> </tr> <tr> <td>Experimenting with initial use cases to familiarize testers with Generative AI</td> </tr> <tr> <td>Managing the evolution of test processes following the full integration of Generative AI</td> </tr> </tbody> </table> </div> <div data-bbox="1196 735 1503 975"> <table border="1"> <thead> <tr> <th>Key phases</th> </tr> </thead> <tbody> <tr> <td>Discovery</td> </tr> <tr> <td>Initiation and Usage Definition</td> </tr> <tr> <td>Exploitation and Iteration</td> </tr> </tbody> </table> </div> </div> 	Activities	Identifying and prioritizing practical use cases	Training testers on Generative AI	Providing testers access to LLMs	Experimenting with initial use cases to familiarize testers with Generative AI	Managing the evolution of test processes following the full integration of Generative AI	Key phases	Discovery	Initiation and Usage Definition	Exploitation and Iteration	<p>GenAI-5.1.4</p>	<p>K1</p>	<p>1</p>
Activities															
Identifying and prioritizing practical use cases															
Training testers on Generative AI															
Providing testers access to LLMs															
Experimenting with initial use cases to familiarize testers with Generative AI															
Managing the evolution of test processes following the full integration of Generative AI															
Key phases															
Discovery															
Initiation and Usage Definition															
Exploitation and Iteration															