文档编号：EA00020_ISTQB_CT_GenAI_EN1.1_CN2.0

认证测试工程师
专业域生成式人工智能测试
模拟卷 A-答案

（大纲 1.1 版）

版本号：EN1.1_CN2.0
发布日期：2026 年 07 月 24 日

国际软件测试认证委员会

中文版的翻译、编辑和出版统一由 ISTQB®授权的 CSTQB®负责

若您对此文档有任何问题，欢迎您扫码添加【官方微信号】反馈。

版权声明

文档责任

国际软件测试认证委员会（ISTQB®）考试工作组对本文件负责。

本文件由 ISTQB®的核心团队维护，该团队成员来自大纲工作组与考试工作组。

致谢

本文件由国际软件测试认证委员会（ISTQB®）的核心团队编写，成员包括：Abbas Ahmad, Alessandro Collino, Bruno Legeard.

核心团队感谢考试工作组评审团队、大纲工作组以及各成员委员会提出的建议与意见。

本文档是属于 ISTQB® 专业域大纲-生成式人工智能测试 1.1 版本的附件，由作为 ISTQB®中国成员国委员会 CSTQB®组织专家团队统一进行了本地化工作，在此感谢参加此次生成式人工智能测试中文本地化工作组专家成员。参加翻译和评审工作的成员有（按姓氏拼音排序）：白宇、陈飞、贺炘（组长、QA 评审）、李健、刘海英、刘晓更、叶岚、张喆、周杨虹。

目录

修订历史

样题 - 使用的题目布局模版:	版本 2.12	日期: 2024 年 11 月 27 日

版本	日期	备注
V1.0	2024/12/30	CT-GenAI 模拟卷 A 1.0 Alpha Release
V1.0	2025/04/17	CT-GenAI 模拟卷 A 1.0 Beta Release
V1.0	2025/06/10	CT-GenAI 模拟卷 A 1.0 Release
EN1.0_CN1.0	2026/04/13	CT-GenAI 模拟卷 v1.0 中文本地化发布
V1.1	2026/04/27	CT-GenAI 模拟卷 v1.1 微小修订
EN1.1_CN2.0	2026/07/24	CT-GenAI 模拟卷 v1.1 中文本地化发布

引言

文档目的

本模拟试卷中的例题、答案以及相关的解析是由领域专家和经验丰富的出题人员编制而成，目的如下：

- 为 ISTQB® 成员委员会和考试委员会在编写考题时提供帮助。
- 为培训机构和考生提供考试题目示例。

这些样题不能在任何正式考试中直接使用。

注意，实际考试可能包含各种各样的题目。这份样题并**未涵盖**所有可能的题型、风格或篇幅，且本样题难度可能高于或低于任何正式考试。

说明

在本文件中，你会看到：

- 答案表中，包含每个正确答案的：
  - K 级别、学习目标、分值
- 答案集，包括所有问题的：
  - 正确答案
  - 每个回答（答案）的详细解析
  - K 级别、学习目标、分值
- 附加的答案集，包含所有题目的（并非所有模拟卷都有此项）
  - 正确答案
  - 每个回答（答案）的详细解析
  - K 级别、学习目标、分值
- *问题包含在另一份文档中*

参考答案

问题编号 (#)	正确答案	LO	K-级别	分值		问题编号 (#)	正确答案	LO	K-级别	分值
1	d	GenAI-1.1.1	K1	1		21	b	GenAI-3.1.3	K2	1
2	c	GenAI-1.1.2	K2	1		22	b	GenAI-3.1.4	K1	1
3	b	GenAI-1.1.2	K2	1		23	d	GenAI-3.2.1	K2	1
4	d	GenAI-1.1.3	K2	1		24	d	GenAI-3.2.2	K2	1
5	b	GenAI-1.1.4	K2	1		25	a	GenAI-3.2.2	K2	1
6	a, e	GenAI-1.2.1	K2	1		26	b	GenAI-3.2.3	K2	1
7	d	GenAI-1.2.2	K2	1		27	c	GenAI-3.3.1	K2	1
8	b	GenAI-2.1.1	K2	1		28	b, d	GenAI-3.4.1	K1	1
9	c	GenAI-2.1.1	K2	1		29	a	GenAI-4.1.1	K2	1
10	b	GenAI-2.1.2	K2	1		30	a	GenAI-4.1.2	K2	1
11	a	GenAI-2.1.3	K2	1		31	c	GenAI-4.1.3	K2	1
12	a	GenAI-2.2.1	K3	2		32	b	GenAI-4.2.1	K2	1
13	b	GenAI-2.2.2	K3	2		33	b	GenAI-4.2.2	K2	1
14	c	GenAI-2.2.3	K3	2		34	d	GenAI-5.1.1	K1	1
15	c	GenAI-2.2.4	K3	2		35	b	GenAI-5.1.2	K2	1
16	a	GenAI-2.2.5	K3	2		36	b	GenAI-5.1.3	K2	1
17	a	GenAI-2.3.1	K2	1		37	a	GenAI-5.1.4	K1	1
18	a	GenAI-2.3.2	K2	1		38	c	GenAI-5.2.1	K2	1
19	c	GenAI-3.1.1	K1	1		39	c	GenAI-5.2.2	K1	1
20	d	GenAI-3.1.2	K3	2		40	a	GenAI-5.2.3	K1	1

答案

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
1	d	解析： • 符号 AI（Symbolic AI）采用基于规则的系统来模拟人类决策，使用符号和逻辑规则表示知识（1B）。 • 传统机器学习（Classical machine learning）采用数据驱动方法，需要数据准备、特征选择与模型训练（2D）。 • 深度学习（Deep learning）利用神经网络从数据中自动学习特征（3A）。 • 生成式 AI（Generative AI）利用深度学习技术，通过从训练数据中学习并模仿模式来创建新数据（4C）。 所以: a) 不正确。 b) 不正确。 c) 不正确。 d) 正确。	GenAI-1.1.1	K1	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
2	c	a) 不正确，因为上下文窗口不控制时序顺序，而是限制可同时处理的文本范围。 b) 不正确，因为上下文窗口影响的是当前输入范围内的内容，而非跨文档引用能力。 c) 正确。当文本长度超过上下文窗口大小时，模型无法同时处理文档的所有部分。当模型处理新词元时，必须有效 “遗忘” 或丢弃超出上下文窗口边界的早期词元。 d) 不正确，因为上下文窗口不决定解析方式，只是限制可同时处理的文本总量。	GenAI-1.1.2	K2	1
3	b	a) 不正确。这描述的是嵌入（embeddings），而非词元化。 b) 正确。词元化是将文本切分成更小的单元（词元 /token），这些单元是自然语言生成任务的基本构建块，使大语言模型能够理解和生成文本。见大纲中 “词元化” 的定义（见附录 D—— 生成式 AI 专用术语）。 c) 不正确。这描述的是大语言模型的一般功能，而非词元化。 d) 不正确。这指的是大语言模型生成文本的方式，而非词元化。	GenAI-1.1.2	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
4	d	解析： i. 不正确。虽然基础大语言模型可以生成测试用例，但在无结构化输入的情况下，它们本身并不 “擅长” 此任务。无结构化输入生成测试用例的描述夸大了其能力 ii. 不正确。创建基于模板的测试脚本需要执行明确指令，这是指令微调大语言模型的任务，而非推理大语言模型。推理大语言模型专注于逻辑推理与问题求解，而非严格遵循模板。 iii. 不正确。指令微调大语言模型旨在遵循结构化提示，而非自主决策。根据反馈对测试进行优先级排序需要推理能力，这超出了它们的能力范围。 iv. 正确。推理大语言模型专门用于多源数据汇总、逻辑推理、问题求解与决策。发现趋势并确定测试优先级，符合其上下文分析的定位。 v. 正确。指令微调大语言模型经过专门训练来遵循指令，包括遵守指定格式、风格和语法规则。生成符合 Gherkin 语法的测试用例非常适合它们。 因此： a) 不正确。 b) 不正确。 c) 不正确。 d) 正确。	GenAI-1.1.3	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
5	b	a) 不正确。视觉语言模型是多模态大语言模型的子集，而非相反 b) 正确。视觉语言模型专门处理视觉与文本数据，属于多模态大语言模型的子集。 c) 不正确。视觉语言模型与多模态大语言模型密切相关，并且同时专注于视觉和文本数据。 d) 不正确。多模态大语言模型与视觉语言模型范围不同，不可互换使用。	GenAI-1.1.4	K2	1
6	a, e	a) 正确。大语言模型可以通过识别需求中的歧义与不一致之处，对需求进行分析和澄清。 b) 不正确。生成完整的应用程序代码并非大语言模型在测试任务中的关键能力。 c) 不正确。大语言模型可通过优化测试脚本、识别设计模式来辅助测试自动化，但在无人监督的情况下无法直接执行测试脚本，也无法完全自动化测试脚本。 d) 不正确。大语言模型无法开展手工探索性测试，因为这一过程需要凭借人的创造力、经验和决策能力，具有直观性与适应性。 e) 正确。大语言模型可以生成多样化的测试数据，包括各种组合与边界值。	GenAI-1.2.1	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
7	d	a) 不正确。AI 聊天机器人最适合用于临时交互，而不是特定的测试任务。 b) 不正确。AI 聊天机器人和由基于大语言模型的测试应用目的不同，功能性和可配置性也不相同。 c) 不正确。基于大语言模型的测试应用重点在于融入测试过程，而非依赖对话式提示。AI 聊天机器人无需集成到测试工具与测试过程中，因为其主要功能是推动临时交互，并非执行特定测试任务。 d) 正确。AI 聊天机器人为临时任务提供对话界面，而基于大语言模型的测试应用则针对特定需求提供定制化解决方案。	GenAI-1.2.2	K2	1
8	b	a) 不正确。上下文提供的是关于测试环境或被测系统的背景信息，并非待分析的具体（输入）数据。此处虽提及性能测试，但列举的是实际数据源，而非背景信息。 b) 正确。该描述罗列了大语言模型执行分析任务时将处理的特定数据源，即测试报告、监控日志和性能基准。 c) 不正确。约束条件概述的是执行任务时的限制因素或特殊注意事项。此描述列举的是实际数据源，而非对分析方法的限制。 d) 不正确。输出格式规定的是大语言模型响应的预期结构与特点。该描述列举的是实际数据源，并未涉及结果应如何呈现。	GenAI-2.1.1	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
9	c	a) 不正确。给定内容并非描述要执行的任务，而是阐述结果的呈现方式。“指令”是告知大语言模型要做什么，而这一行内容是在告知大语言模型如何对输出进行格式编排。 b) 不正确。给定内容并非对分析过程设置限制条件，而是明确所需的呈现格式。“约束” 可能会涉及 “排除表面性问题” 这类内容，而此内容围绕的是输出格式。 c) 正确。给定内容与大纲中 “输出格式” 的定义相符，其作用在于引导大语言模型对输出进行格式编排。 d) 不正确。给定内容未提供如需求规格说明书背景等信息，它关注的是输出格式。	GenAI-2.1.1	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
10	b	a) 不正确。提示词链实际上是将任务分解为一系列顺序的提示，并非 “提供示例”。少样本提示是通过提供示例来进行引导的技术，并非 “将任务分解为子任务”。元提示依靠大语言模型自行修正提示，而非测试人员“手动优化提示”。 b) 正确。少样本提示借助示例提供指导，提示词链将任务分解为多个中间步骤（多个提示），元提示则利用人工智能迭代优化自身提示词。 c) 不正确。元提示的核心是让大语言模型自动优化提示词，并非 “将任务分解为多个步骤”。提示词链的特点是逐步分解任务，并非如选项所述“使用示例”。少样本提示是向模型提供示例的方法，重点并非 “手动优化提示词”。 d) 不正确。提示词链的定义是逐步分解任务，并非 “在无示例情况下提供指导”。元提示依赖大语言模型生成或优化提示，并非仅依靠测试人员定义的措辞。	GenAI-2.1.2	K2	1
11	a	a) 正确。在整个交互过程中，系统提示词始终保持不变，它为大语言模型的响应方式构建了基本框架。 b) 不正确。这所描述的是用户提示词的功能，并非系统提示词的功能。 c) 不正确。系统提示词不会动态调整，它保持固定不变。 d) 不正确。系统提示词是隐藏的，并不包含用户可见的输入内容。	GenAI-2.1.3	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
12	a	解析： i. 正确。因为它确保整个过程从依据需求生成的测试条件开始，这与大纲所描述的 “利用生成式 AI 进行测试分析，需基于测试依据（如需求）来生成测试条件”相符。 ii. 正确，因为它明确了验收准则（测试条件）对于优化大语言模型生成结果的重要性，并遵循大纲的指导：即当给予适当的上下文时，大语言模型能够依据风险等级对测试条件进行优先级排序。 iii. 正确，因为它要求大语言模型进行覆盖范围分析，以覆盖需求的所有方面，这与大纲中“大语言模型能够进行覆盖分析以判定测试依据的所有方面是否均得到覆盖”的表述相符。 iv. 不正确，因为依赖最少的输入而没有指导不符合提示词链技术，也无法保证有效性。这个选项试图一步完成所有工作，而不是将任务拆解。 v. 不正确，因为缺陷识别并非生成优先级测试条件和识别覆盖缺口这一测试目标的核心内容。题目场景明确说明需求是稳定的，并且已经全面评审过缺陷（通常包括模糊和不一致）。 因此： a) 正确。 b) 不正确。 c) 不正确。 d) 不正确。	GenAI-2.2.1	K3	2

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
13	b	a) 不正确。虽然它指定使用预定义示例，但并未明确要求使用“Given-When-Then”语法，而这对于 Gherkin 风格的测试用例至关重要。它还表明依赖于模糊的最佳实践来定义约束条件，并且没有明确确保与验收准则保持一致。 b) 正确。全面且利用预定义示例来指导 LLM。 c) 不正确。缺乏对使用预定义示例的重视，并且表明依赖于模糊的最佳实践来定义指令。 d) 不正确。只关注了边界情况，但忽略了全面的覆盖范围和使用示例进行指导。	GenAI-2.2.2	K3	2
14	c	a) 不正确。此选项仅涉及问题归类和交叉核对，却遗漏了诸如区分测试结果等其他关键步骤。 b) 不正确。虽然提高了角色清晰度，但既未对指令进行扩充，也未涉及结构化步骤。 c) 正确。该选项对指令进行了扩展，纳入了所有关键的结构化分析步骤。区分预期结果与实际结果，能有效找出两者的差异；对问题进行归类，有助于更合理地确定优先级，并减少重复内容；突出显示差异，则能让人将注意力聚焦在最重要的发现上。 d) 不正确。引入了无关的约束，导致提示与任务要求不符。	GenAI-2.2.3	K3	2

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
15	c	a) 不正确。虽然明确了角色，但没有增添任何能提升所生成指标准确性、可操作性或可解读性的具体指令。 b) 不正确。添加的风险分析任务可能会让模型偏离核心度量，而且依旧没有提供能让利益相关方清晰理解结果的机制。 c) 正确。使用通俗易懂的语言摘要扩展输出格式，解释度量并规划后续步骤，可以直接帮助利益相关方理解和采取行动（详见大纲中的 2.2.4 节：“强化的测试度量可视化与报告”）。 d) 不正确。只是重复了现有的一个约束条件，对于大语言模型该如何实现让利益相关方易于解读，并未给出具体指引。	GenAI-2.2.4	K3	2
16	a	a) 正确。少样本提示在这种情况下非常理想，因为它允许提供一些示例（现有测试用例及其输入和预期结果）来引导大语言模型。通过展示如何运用这些转换规则生成新的测试用例，少样本提示有助于大语言模型理解并复刻该流程，进而生成更多测试用例。 b) 不正确。虽然可以使用提示词链，但对于这样一个极为简单直接的任务而言，它可能会无端增加复杂度。 c) 不正确。尽管元提示也可采用，然而在应对这一简单直接的任务时，它并不像少样本提示那般直接且具针对性。 d) 不正确。零样本提示不会有效，因为它无法借助现有测试用例作为示例。	GenAI-2.2.5	K3	2

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
17	a	a) 正确。多样性可确保全面涵盖边界用例，测试执行成功率用于评估 API 测试脚本的可靠性（见大纲 2.3.1 节表格中 “多样性” 度量的描述和示例）。 b) 不正确。虽然准确性和完整性很重要，但主要依赖时间效率并不能全面评估覆盖率或测试执行的可靠性。 c) 不正确。精确度和上下文契合度固然有价值，但无法解决多样性问题，也不能全面评估 API 测试脚本的测试执行成功率。 d) 不正确。相关性和上下文契合度至关重要，但如果不考虑测试执行成功率或准确性等度量，就会忽略评估的关键方面。	GenAI-2.3.1	K2	1
18	a	a) 正确。输出分析旨在检查大语言模型的输出是否存在不准确或不一致的情况。通过对与需求相悖的错误预期结果进行分类，能够揭示提示词误导大语言模型的原因，为优化提示词提供具体思路。 b) 不正确。提示词的 A/B 测试更适用于比较不同版本的提示词，而非诊断错误预期结果产生的原因。 c) 不正确。尽管调整提示词的长度和具体程度，可通过增减上下文来提升回复质量，但无法直接找出预期结果错误产生的根源。 d) 不正确。虽然借助从测试人员处收集到的关于生成输出的实用性与清晰度的见解，有助于优化提示词以更好地满足实际测试需求，但这并不能直接解释错误预期结果产生的原因。	GenAI-2.3.2	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
19	c	a) 不正确。此描述针对的是推理错误，并非幻觉。 b) 不正确。这说的是人工智能输出中的偏差，并非幻觉。 c) 正确。当大语言模型生成的内容与事实不符或与给定任务无关时，便出现了幻觉。详见大纲 3.1.1 节。 d) 不正确。这指的是由于训练数据代表性不足所导致的偏差，并非幻觉。	GenAI-3.1.1	K1	1
20	d	a) 不正确。项目简报中明确提到了购物车管理，所以这个测试用例是相关的。 b) 不正确。项目简报明确提及了折扣码应用，该测试用例与之相关。 c) 不正确。项目简报明确提到了订单确认邮件，此测试用例有效。 d) 正确。项目简报未提及心愿单管理，所以这个测试用例最有可能是 “幻觉”生成的。	GenAI-3.1.2	K3	2
21	b	a) 不正确。针对测试任务微调大语言模型的工作量，与在目标数据集上对模型进行额外训练相关，以便其学习执行这些测试任务所需的特定领域知识和细微差别。使用清晰、结构化的输入数据格式并不会影响这一工作量。 b) 正确。当输入数据以清晰、结构化的方式呈现时，潜在误解会降至最低。歧义往往源于不清晰且结构欠佳的数据。 c) 不正确。上下文相关性更多依赖于提供恰当的上下文信息，而非仅仅依靠清晰、结构化的输入数据格式。 e) 不正确。使用清晰、结构化的输入数据格式并不会提高大语言模型生成输出的创造性。特别是，此类格式并不鼓励新颖的回复，因为它要求大语言模型按照指定格式生成回复内容。	GenAI-3.1.3	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
22	b	a) 不正确。学习率是训练开始前选定的设置，用于控制神经网络的学习过程，它决定训练中模型权重的变化幅度，与推理时输出的可变性无关。提高学习率能加快训练收敛速度，但不影响推理时输出的可变性。 b) 正确。“温度” 是一个参数，在推理时通过作用于概率分布来控制输出的随机性。降低温度参数可减少随机性，使输出更趋一致。 c) 和 d) 不正确。设置随机种子值可提高结果的可重复性，但本身并不缩小推理时的概率分布，而且对其数值高低并无影响。	GenAI-3.1.4	K1	1
23	d	a) 不正确。这属于隐私问题，因为“非故意数据暴露”指的是生成式 AI 在输出时意外泄露敏感信息。 b) 不正确。这同样是隐私问题，因其涉及对敏感数据存储与处理缺乏管控。 c) 不正确。这是与未遵守《通用数据保护条例》（GDPR）等法规相关的隐私问题，会带来法律风险。 d) 正确。只要大语言模型训练时没用真实敏感数据，那么它在生成合成测试数据产生 “幻觉” 时，就不会暴露真实敏感数据。这种情况下的 “幻觉” 内容是纯合成的，基于模型训练时学到的模式和结构。大语言模型可能无意中生成与真实敏感数据相符的合成测试数据（这确实值得关注），但这并非暴露真实敏感数据。不过，这种无意生成的可能性极低。	GenAI-3.2.1	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
24	d	a) 不正确。恶意代码生成是指 “在使用过程中操控大语言模型生成后门程序（比如外部命令调用）”，并非向训练数据集中注入虚假数据。 b) 不正确。上下文操控是指 “发送意在提取机密训练数据的请求”，并非向训练数据集中注入虚假数据。 c) 不正确。请求篡改是指在运行使用阶段 “引入会干扰 AI 输出的数据”，并非向训练数据集中注入虚假数据。 d) 正确。根据大纲 3.2.2 节，数据投毒是指 “操控训练数据”，具体例子为 “在给 AI 生成的测试报告结果评分时提供虚假评估”。本题所描述的正是这种攻击方式：攻击者将伪造的测试（执行）结果注入训练数据集，直接操控训练数据，破坏大语言模型准确推荐最优测试覆盖策略的能力。	GenAI-3.2.2	K2	1
25	a	考虑到： • 上下文操控指发送旨在提取机密训练数据的请求。(1C) • 请求篡改指引入会扰乱大语言模型输出的图像。(2D) • 数据投毒指通过在训练中使用虚假评估来操控推荐内容。(3A) • 指在使用过程中操控大语言模型生成后门（例如外部命令调用）。(4B) 因此： a) 正确。 b) 不正确。 c) 不正确。 f) 不正确。	GenAI-3.2.2	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
26	b	a) 不正确。尽管通过对比多个大语言模型来评估输出是一种有效的做法，但在此情形下，它侧重于从正确性角度考量输出质量，并未解决数据隐私风险问题。 b) 正确。匿名化是降低数据隐私风险的有效策略。 c) 不正确。对敏感数据的无限制访问会加大敏感数据泄露的风险。 d) 不正确。禁用敏感数据的加密会削弱安全性，增加敏感数据遭窃取的风险。	GenAI-3.2.3	K2	1
27	c	a) 不正确。图像生成远比文本生成消耗更多能源并产生更多二氧化碳排放。除非指定了关键因素（如能源来源差异），否则能源消耗与二氧化碳排放通常呈正相关。 b) 不正确。恰恰相反，生成式 AI 驱动的搜索比传统网络搜索消耗的能源要多得多。 c) 正确。图像生成比文本生成需要多得多的计算资源，这使得它能源消耗程度高得多。 d) 不正确。数百万用户进行文本生成所带来的累积影响不容小觑。	GenAI-3.3.1	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
28	b, d	a) 不正确。ISO/IEC 25010:2023 是一项标准（本大纲未提及，但 CTFL 大纲中有提及），该标准定义了与信息通信技术（ICT）/ 软件产品质量属性相关的产品质量模型。它并未涉及生成式 AI 在软件测试中的应用。 b) 正确。ISO/IEC 23053:2022 是一项标准（在大纲 3.4.1 节中有提及），它为使用生成式 AI 进行测试时的数据质量、透明度和容错能力提供了一个框架。 c) 不正确。ISO/IEC/IEEE 29119 - 2:2021 是该标准的一部分（卷）（本大纲未提及，但 CTFL 大纲中有提及），主要涉及测试过程。它并未涉及生成式 AI 在软件测试中的应用。 d) 正确。ISO/IEC 42001:2023 是一项标准（在大纲 3.4.1 节中有提及），它规定了组织内管理基于 AI 系统的要求，为生成式 AI 在软件测试中的应用提供了最佳实践，并有助于提升一致性和可靠性。 e) 不正确。ISO/IEC/IEEE 29119 - 3:2021 是该标准的一部分（卷）（本大纲未提及，但 CTFL 大纲中有提及），主要涉及测试文档。它并未涉及生成式 AI 在软件测试中的应用。	GenAI-3.4.1	K1	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
29	a	a) 正确。后端负责从关系型数据库和向量数据库中检索数据，将其与用户输入相结合，进而为大语言模型准备定制化的提示词。 b) 不正确。前端作为用户提交输入的界面，并不负责为大语言模型准备提示词。 c) 不正确。认证组件的作用是确保访问安全，与数据检索或提示词准备并无关联。 d) 不正确。后处理组件对大语言模型生成的输出进行优化，并不参与提示词准备工作。	GenAI-4.1.1	K2	1
30	a	a) 正确。检索增强生成（RAG）会自动检索相关规格说明以及历史测试用例片段，然后将它们输入到大语言模型中，接着大语言模型会生成准确且具有上下文感知能力的测试用例。 b) 不正确。检索增强生成（RAG）在从向量数据库中检索特定的目标信息时能发挥更大作用，而不是检索整个数据集。 c) 不正确。手动审查并优化查询会带来不必要的工作量，这与检索增强生成（RAG）自动化检索以及在大语言模型生成回复时动态使用检索到的数据这一设计理念相悖。 d) 不正确。依赖大语言模型的内部数据，就忽视了检索增强生成（RAG）通过整合外部最新信息来强化回复内容的核心能力。	GenAI-4.1.2	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
31	c	a) 不正确。自主和半自主式智能体可以采用单一智能体和/或多智能体系统的形式实现，以提高测试过程的效率和质量。然而，这些智能体带来的关键提升在于，它们能够在自主性与人工监督之间取得平衡，从而提高测试过程的效率和质量。 b) 不正确。虽然自主和半自主智能体整合了验证机制以确保质量，但这些验证机制在设计上同样追求高效。采用自动化检查实际上是为了简化测试过程，在不牺牲效率的情况下提升质量（反之，效率也会相应提升）。 c) 正确。自主智能体在测试过程中，通过最少的人工干预执行测试任务，并采用精心设计的高效自动化检查来提高效率。半自主智能体则包含策略性的人工监督，以此确保测试结果的质量。这两类智能体能够借助自身在不同程度人机交互下运行的能力，实现一种兼顾效率与质量的平衡方法。 d) 不正确。完全取消验证既不现实也不可取。即使使用自主和半自主式大语言模型驱动的智能体这类先进技术，验证仍然至关重要。	GenAI-4.1.3	K2	1
32	b	a) 不正确。这是对微调的正确描述，微调通过面向特定任务的训练来增强模型在特定领域的性能。 b) 正确。这个陈述是不正确的，因为微调并非取代模型的通用知识，且过拟合始终是潜在难题。 c) 不正确。这是一个正确的描述，因为微调确实涉及使用目标数据集进行参数调整。 d) 不正确。这是对微调相关挑战的正确描述，因为高质量的数据集对于有效适配至关重要。	GenAI-4.2.1	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
33	b	a) 不正确。大语言模型运维（LLMOps）侧重于管理大语言模型（LLM），而非阻止其使用。 b) 正确。大语言模型运维在大语言模型的生存周期内对其进行管理，解决测试中的隐私、安全和成本问题。 c) 不正确。大语言模型运维适用于各种应用场景，而不仅仅是基于聊天机器人的解决方案。 d) 不正确。大语言模型运维的目标并非完全自动化所有测试任务也并非消除人工监督。	GenAI-4.2.2	K2	1
34	d	a) 不正确。使用未经批准的生成式 AI 工具无法确保组织数据政策得以执行。 b) 不正确。影子 AI 会引发与许可协议不明确相关的风险，进而可能导致知识产权纠纷。 c) 不正确。影子 AI 增加而非降低了知识产权纠纷的风险。 d) 正确。未经批准的 AI 工具往往缺少完善的安全措施，这就加大了数据泄露或未经授权访问的风险。	GenAI-5.1.1	K1	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
35	b	a) 不正确。尽管针对特定大语言模型的认证可能有用（且未来可能会更加普及），但培训计划应旨在确保测试团队成员具备有效使用生成式 AI 工具所需的技术技能。 b) 正确。选择与现有测试基础设施和可扩展性要求兼容的大语言模型，是生成式 AI 策略的一个关键要素。测试工具和测试环境是测试基础设施的基本要素。 c) 不正确。对于借助生成式 AI 获得可靠结果而言，数据质量以及结构化、安全的输入至关重要。需要有适量的（依据测试目标）高质量输入数据，而非越多越好。 d) 不正确。第 5.1.2 节指出，用于软件测试的生成式 AI 策略必须收集特定任务的度量，以评估大语言模型输出的有效性。第 2.3.1 节中列出的度量（例如，相关性、执行成功率、时间效率）是为测试任务定制的，而不是直接从经典的监督式机器学习中借鉴过来的。	GenAI-5.1.2	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
36	b	a) 不正确。关键选择标准包括根据组织自身的基准，使用大纲中提及的度量（见第 5.1.3 节）来评估模型执行测试任务的性能。如果测试任务涉及代码生成，则指定的标准可能适用。然而，它并不适用于其他类型的测试任务。 b) 正确。关键选择标准包括评估周期性成本（见大纲第 5.1.3 节），此答案提及的是周期性成本的一个典型例子。 c) 不正确。参照公开可用的社区基准评估大语言模型，以确保其与之完全兼容，并不是大纲（见第 5.1.3 节）中提及的，为特定测试任务选择合适大语言模型的关键标准之一。 d) 不正确。关键选择标准包括评估周期性成本（见大纲第 5.1.3 节），但本答案提及的是一次性成本的一个典型示例。	GenAI-5.1.3	K2	1

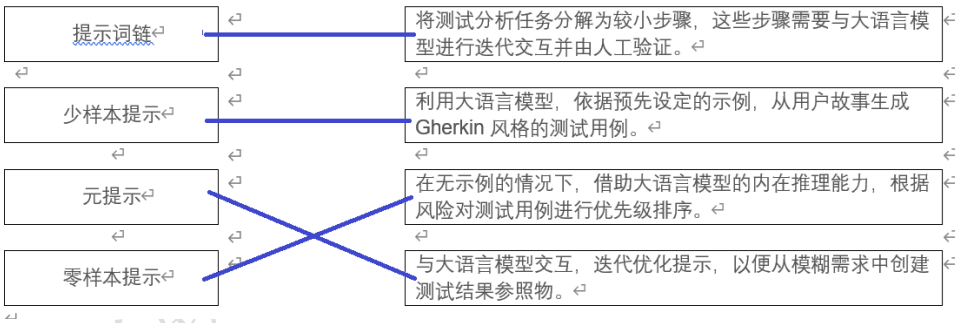
答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
37	a	大纲阐述了以下三个关键阶段： • 探索。 • 启动与使用定义。 • 应用与迭代。 所以： a) 正确：明确提及了这三个关键阶段。 b) 不正确：未明确提及这三个阶段中的任何一个（仅提及了与这些阶段相关的部分目标和 / 或活动）。 c) 不正确：未明确提及这三个阶段中的任何一个（仅提及了与这些阶段相关的部分目标和 / 或活动）。 d) 不正确：未明确提及这三个阶段中的任何一个（仅提及了与这些阶段相关的部分目标和 / 或活动）。	GenAI-5.1.4	K1	1

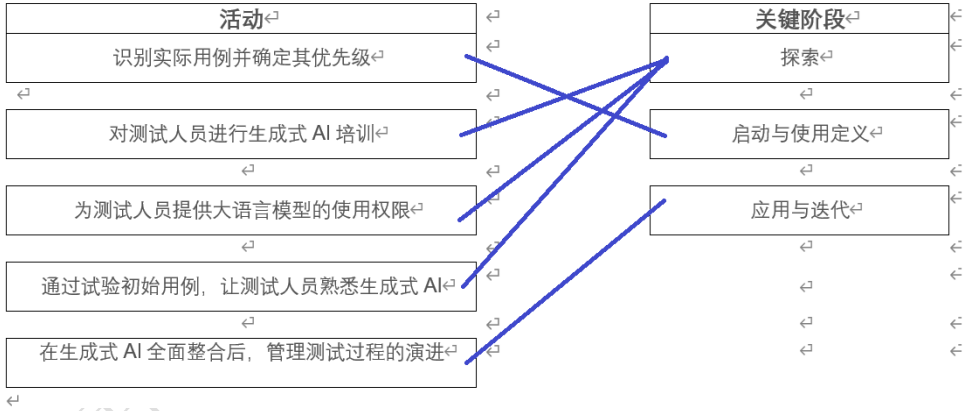
答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
38	c	a) 不正确。大语言模型中的幻觉现象是当前 AI 技术所固有的难题，测试人员无法防止幻觉和推理错误的出现。相反，测试人员在运用生成式 AI 进行测试时，应当能够（识别并）减轻幻觉、推理错误（以及偏差）所带来的风险。 b) 不正确。虽然精通测试自动化很有价值，但它并未专门针对将生成式 AI 整合到测试过程中这一情况。 c) 正确。大纲将 “评估大语言模型能力” 列为一项关键能力。这必然涉及对候选模型进行比较，并挑选出能够针对特定测试任务进行适配或微调的模型。这是测试人员有效使用大语言模型所需知识 / 技能的一个直接示例。 d) 不正确。这指的是开发大语言模型所需的技能，比如人工智能研究人员和数据科学家应具备的技能，而非使用生成式 AI 执行测试任务的测试人员所需的技能。这些测试人员关注的是确保在与大语言模型交互时直接使用的测试数据的质量。	GenAI-5.2.1	K2	1

答案(#)	正确答案	解释 / 理由	学习目标(LO)	K-级别	分值
39	c	a) 不正确。虽然带有实践操作的外部专家课程可能有益，但主要依赖此类课程会削弱内部实践和社区建设的重要性。此外，不推荐以"大爆炸"的方式将 AI 融入所有日常测试任务。 b) 不正确。虽然尝试探索是学习过程的一部分，但毫无章法的独立尝试可能无法持续有效地提升技能。 c) 正确。通过有组织的活动、同伴学习以及知识共享社群，着重培养实践技能，这才是值得推荐的方法。 d) 不正确。过度依赖外部专家的理论课程，会降低通过组织内部交流来培养实践技能与专业知识的重要性。	GenAI-5.2.2	K1	1
40	a	a) 正确。测试人员会演变为 AI 辅助测试专家，负责优化提示词并验证 AI 的输出结果。 b) 不正确。测试经理的职责拓展为涵盖制定基于人工智能的测试策略、开展基于人工智能的风险管理以及监督基于人工智能的测试流程。所以，他们的重点依旧是测试管理，而非钻研生成式人工智能技术（技术层面复杂的）内部运行原理。 c) 不正确。测试人员的工作重心并非转移到监管基于人工智能的测试过程上，这项监管职责仍归测试经理。 d) 不正确。测试经理必须平衡人类与 AI 的能力，以实现高效的测试结果。	GenAI-5.2.3	K1	1

附录：附加题答案

<table>
<tr><th>问题编号 (#)</th><th>正确答案</th><th>解析 / 理由</th><th>学习目标 (LO)</th><th>K-级别</th><th>分值</th></tr>
<tr><td>A1</td><td>不适用</td><td>检索增强生成（RAG）过程可总结如下：
1. 将文档拆分为片段，以便进行有针对性的检索。
2. 清理和处理片段，确保一致性。
3. 将片段编码为嵌入向量，并存储在向量数据库中。
4. 在处理用户查询时检索相关片段。
5. 结合检索到的数据与大语言模型的知识生成回复。
因此，正确答案为：
<table><tr><th colspan="2">答案</th></tr><tr><td>首先</td><td>将大文档拆分为较小的文档块</td></tr><tr><td></td><td>清理并处理文档块</td></tr><tr><td></td><td>将嵌入内容存储于向量数据库中</td></tr><tr><td></td><td>基于语义相似度检索相关文档块</td></tr><tr><td>最后</td><td>使用检索到的文档块和大语言模型生成响应</td></tr></table></td><td>GenAI-4.1.2</td><td>K2</td><td>1</td></tr>
</table>

问题编号 (#)	正确答案	解析 / 理由	学习目标 (LO)	K-级别	分值
A2	不适用	考虑到： • 提示词链 – 右侧第一项直接涉及为完成测试分析任务而对大语言模型（LLM）进行迭代使用（2.2.1 节）。 • 少样本提示 –右侧第二项明确表明通过示例与大语言模型生成的测试用例相关联（2.2.2b 节）。 • 元提示（C） – 右侧第四项着重强调了为优化提示词与大语言模型展开协作（2.1.2 节）。 • 零样本提示（D） – 右侧第三项突出了大语言模型基于内在推理进行优先级排序（2.1.2 节）。 因此，正确答案是：  	GenAI-2.1.2	K2	1

问题编号 (#)	正确答案	解析 / 理由	学习目标 (LO)	K-级别	分值
A3	不适用	考虑到： • 探索阶段：此阶段活动涵盖对测试团队开展生成式 AI 概念培训、为其提供大语言模型/ 小语言模型的使用权限，以及通过初步用例试验，助力测试人员熟悉生成式 AI 并建立信心。 • 启动与使用定义阶段：该阶段重点在于识别生成式 AI 在软件测试中的实际用例，并对其进行优先级排序。 • 应用与迭代阶段：持续监测生成式 AI 在软件测试及相关工具上的进展情况，同时对变革进行评估与管理，以保障可持续效益和可扩展性。 因此，正确答案是：  	GenAI-5.1.4	K1	1