

认证测试工程师 专业域生成式人工智能测试 模拟题-问题

(大纲1.0版)

版本号: EN1.0_CN1.1

发布日期: 2026年04月13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

英文版权声明

版权声明©国际软件测试认证委员会（以下简称“ISTQB®”）。

ISTQB®是国际软件测试资格委员会的注册商标。

保留所有权利。

作者在此将版权转让给 ISTQB®。作者(作为当前版权持有人)和 ISTQB®(作为未来版权持有人)同意以下使用条件：

本文件的摘要部分，如果非商业用途且来源已获确认，则可以复制。

任何经认可的培训机构可在其培训课程中使用这一模拟试卷，条件是承认作者和 ISTQB®是模拟试卷的来源和版权所有人，并且只有在收到 ISTQB®认可的成员委员会对培训材料的正式认可之后，才能为这一培训课程做广告。

任何个人或团体都可以在文章和书籍中使用此模拟试卷，前提是作者和 ISTQB®被确认为模拟试卷的来源和版权所有人。

在没有首次得到 ISTQB®的书面批准之前，禁止任何其他使用该模拟试卷。

任何 ISTQB®认可的成员委员会都可以翻译本模拟试卷，只要他们将上述版权声明复制到模拟试卷的翻译版本中。

中文版权声明

版权标志©国际软件测试认证委员会中国分会（以下简称“CSTQB®”）。

在认可 ISTQB®/CSTQB®为本文档所有者的前提下，可以完整复制本文档或提取摘录，且必须指明出处，但不能用作商业用途。

文件说明

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。

中国软件测试认证委员会 (CSTQB®)

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 1.0 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

版权声明.....	2
文件说明.....	3
致谢	4
目录	5
修订历史.....	6
引言	7
文档目的.....	7
说明.....	7
问题	8
附录：附加题.....	24

中国软件测试认证委员会 (CSTQB®)

修订历史

模拟试卷-使用的问题分布模板： V2.11 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
EN1.0_CN1.1	2026/04/13	CT-GenAI 模拟卷 v1.0 部分修订

中国软件测试认证委员会 (CSTQB®)

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：

- 问题1，每题包含：
 - 题干所需的任何场景设定
 - 分值
 - 选项（答案）集
- 附加题目，（并非所有模拟卷都有此项），每题包含：
 - 题干所需的任何情景设定
 - 分值
 - 选项（答案）
- 答案及理由解析则包含在另一份单独文档中

¹ 在本模拟卷中，题目是按照其对应的学习目标（LO）进行排序的；但实际考试中可能并非如此。

问题

1. 将每种人工智能技术类型（1-4）与其正确描述（A-D）进行匹配：

1. 符号 AI
 2. 传统机器学习
 3. 深度学习
 4. 生成式 AI
-
- A. 使用神经网络从数据中自动学习特征
 - B. 使用基于规则的系统来模拟人类决策。
 - C. 使用深度学习，通过从训练数据中的学习来生成新数据。
 - D. 使用数据驱动方法，需要人工进行特征选择。

A	1D, 2B, 3A, 4C
B	1D, 2C, 3B, 4A
C	1C, 2B, 3D, 4A
D	1B, 2D, 3A, 4C

选择一个选项。

2. 在大语言模型领域，以下哪个选项最好地解释了为什么上下文窗口限制会影响大语言模型的文本处理能力？

A	因为上下文窗口限制了时间序列处理序列，导致大语言模型在长文本分析中无法保持时间顺序上的一致性。
B	因为上下文窗口会阻碍交叉引用能力，限制大语言模型同时关联不同文档来源信息的能力。
C	因为上下文窗口会迫使大语言模型丢弃较早的信息，而这些信息可能包含理解后续内容所需的相关细节。
D	因为上下文窗口会限制解析粒度级别，阻碍大语言模型在字符级和文档级分析之间切换。

选择一个选项。

3. 以下哪项陈述**最准确**地描述了大语言模型文本处理中的词元化（Tokenization）

A	词元化将词元转换为高维向量以捕捉其含义。
B	词元化构建用于理解和生成文本的基本单元。
C	词元化利用神经网络生成符合上下文的回复。
D	词元化根据学习到的关系预测序列中的下一个词元。

选择一个选项。

4. 在软件测试领域中，关于基础大语言模型、指令微调大语言模型和推理大语言模型的以下陈述（i - v），哪些是正确的？

- i. 基础大语言模型擅长在无结构化输入的情况下，从高级别需求中生成测试用例。
- ii. 推理大语言模型擅长创建严格遵循组织预定义模板的测试脚本。
- iii. 指令微调大语言模型擅长根据实时用户反馈自主确定测试执行的优先级。
- iv. 推理大语言模型擅长从缺陷报告中汇总数据，以发现趋势并确定测试工作的优先级。
- v. 指令微调大语言模型擅长生成符合 Gherkin 语法的测试用例。

A	i, ii, 和 iii
B	ii, iii, 和 iv
C	i, ii, 和 v
D	iv, 和 v

选择一个选项。

5. 以下哪项陈述最准确地描述了多模态大语言模型与视觉语言模型之间的关系？

A	多模态大语言模型是视觉语言模型的一个子集，用于处理多种输入。
B	视觉语言模型是多模态大语言模型的一个子集，专注于处理视觉和文本数据。
C	视觉语言模型与多模态大语言模型无关，仅专注于用户界面。
D	多模态大语言模型和视觉语言模型是可互换的术语。

选择一个选项。

6. 以下哪两项是大语言模型在测试任务中的关键能力？

A	识别需求中的歧义与不一致之处。
B	生成可直接部署的完整应用程序代码。
C	在无人干预下自动执行所有测试脚本。
D	对软件应用程序开展探索性测试。
E	创建包含多种组合与边界值的多样化测试数据。

选择两个选项。

7. 以下哪一项陈述**最好的**阐释在软件测试场景中，AI聊天机器人与基于大语言模型的测试应用之间的差异？

A	AI聊天机器人更适合执行特定的测试任务，而基于大语言模型的测试应用则更侧重于临时交互。
B	AI聊天机器人和基于大语言模型的测试应用都被设计为执行相同的任务，且在配置上没有任何差异。
C	基于大语言模型的测试应用依赖对话式提示，而 AI 聊天机器人则需要与测试工具和测试过程进行集成。
D	AI聊天机器人为临时测试任务提供对话式界面，而基于大语言模型的测试应用则为特定测试任务提供定制化解决方案。

选择一个选项。

8. 测试人员正在查看一个结构化提示词，该提示词用于借助大语言模型辅助进行性能测试分析。该提示词其中一个部分的内容为：“性能测试工具生成的测试报告、使用高峰期的系统监控日志，以及过往版本的应用性能基准”。

在六部分提示词结构中，该描述**最有可能**出现在哪一部分？

A	上下文。
B	输入数据。
C	约束。
D	输出格式。

选择一个选项。

9. 一名测试人员希望借助大语言模型分析需求规格说明中的潜在缺陷。在测试人员使用的结构化提示词中，有一行内容如下：“潜在缺陷必须以 Markdown 表格形式提供，表格需包含以下列：ID、需求引用、缺陷类型、描述、严重程度”。

这个描述**最有可能**出现在六部分提示词结构的哪一个组成部分？

A	指令。
B	约束。
C	输出格式。
D	上下文。

选择一个选项。

10. 下列哪项最能区分提示词链、少样本提示和元提示技术？

A	提示词链着重于给出示例，少样本提示是把任务拆解成子任务，元提示则是手动优化提示词。
B	少样本提示通过示例来提供指引，提示词链将任务分解为多个提示词，元提示则让模型以迭代的方式优化自身的提示词。
C	元提示强调把任务细化为多个步骤，提示词链运用示例，少样本提示聚焦于手动对提示词进行优化。
D	提示词链在不借助示例的情况下提供指引，少样本提示借助示例提供指引，元提示依赖测试人员定义的提示词。

选择一个选项。

11. 在与大语言模型的交互中，系统提示词的主要功能是什么？

A	为大语言模型在整个对话中的行为提供一个框架。
B	向大语言模型提供用户的具体问题或指令。
C	根据每次用户交互动态调整，并设定对话上下文。
D	包含用户可见的输入并设定对话规则。

选择一个选项。

12. 你负责将以下测试方法应用于一个新项目的一系列稳定需求：生成测试条件、根据风险等级进行优先级排序，并识别潜在覆盖缺口。这些需求已通过全面的缺陷评审。

为有效运用生成式AI，通过提示词链技术来实施该测试方法，你应遵循以下哪一个步骤序列 (i-v)？

- i. 将需求提交给大语言模型，提示其基于需求生成测试条件。
- ii. 向大语言模型提供测试条件，确保其理解优先级排序的上下文，并提示其据此排序。
- iii. 向大语言模型提供排序后的测试条件，提示其分析这些测试条件，判定是否覆盖了需求的所有方面。
- iv. 向大语言模型提交需求，并提示其生成已按优先级排序且涵盖需求所有方面的测试条件。
- v. 向大语言模型提交需求，并提示其检测需求中的不一致和歧义。

A	i, ii, 和 iii
B	iv, 和 ii
C	i, iii, 和 v
D	v, 和 iv

选择一个选项。

13. 考虑运用少样本结构化提示技术，为以下用户故事和验收准则生成 Gherkin 风格（即基于场景的）测试用例：

- 用户故事：“作为用户，我想重置密码，以便在我忘记密码时能够重新获得账户访问权限。”
- 验收准则：“当用户提交了已注册的电子邮件地址后，他们会收到一封密码重置邮件。”

你可以参考预定义的示例，包括用户故事、验收准则和 Gherkin 风格的测试用例。你的任务是创建一个提示词，指导大语言模型生成与上述用户故事验收准则相符的准确测试用例。

以下哪一个提示词**最适合**此任务？

A	提示词 A 角色：作为一名测试分析师。 上下文：你正在测试密码重置功能。 指令：参考以下预定义示例，为用户故事和验收准则生成 Gherkin 风格的测试用例： << 预定义示例 >>。
---	--

	<p>输入数据：<<< 用户故事 >>> 和 <<< 验收准则 >>>。 约束：依据最佳实践创建测试用例。 输出格式：生成带有期望结果的测试用例</p>
B	<p>提示词 A 角色：作为一名测试分析师。 上下文：你正在测试密码重置功能。 指令：参考以下预定义示例，为用户故事和验收准则生成 Gherkin 风格的测试用例： << 预定义示例 >>。 输入数据：<<< 用户故事 >>> 和 <<< 验收准则 >>>。 约束：依据最佳实践创建测试用例。 输出格式：生成带有期望结果的测试用例</p>
C	<p>提示词 C 角色：作为一名测试分析师。 上下文：你正在测试密码重置功能。 指令：为用户故事和验收准则生成 Gherkin 风格的测试用例。依靠最佳实践来创建测试用例。 输入数据：<<< 用户故事 >>> 和 <<< 验收准则 >>>。 约束：使用“Given-When-Then”语法，并确保与验收准则保持一致。 输出格式：遵循给定的 Gherkin 风格测试用例格式。</p>
D	<p>提示词 D 角色：作为一名测试分析师。 上下文：你正在测试密码重置功能。 指令：为用户故事和验收准则生成至少两个 Gherkin 风格的测试用例。专注于边缘用例。 输入数据：<<< 用户故事 >>> 和 <<< 验收准则 >>>。 约束：确保所有测试用例都遵循“Given-When-Then”语法。 输出格式：遵循给定的 Gherkin 风格测试用例格式。</p>

选择一个选项

14. 你的任务是应用结构化提示来分析回归测试结果。以下是该提示词的初稿：

角色：作为一名测试分析师。

上下文：分析最近一次测试执行周期的原始回归测试结果。

指令：识别测试结果中的差异之处。

输入数据：使用附带的包含原始测试结果的文件。

约束：使用已知异常列表进行交叉核对。

输出格式：以表格形式提供差异列表。

现要求你完善这个提示词。就全面的回归测试报告分析而言，以下哪项改进最符合结构化提示工程的最佳实践？

A	增加一个步骤，对相似问题进行归类，并将所得结果与已知异常列表做交叉核对。
B	明确该角色是一名专注于提供可执行见解的回归测试分析师。
C	扩充指令内容，涵盖区分预期结果与实际结果、对问题进行归类以及突出显示差异。
D	在约束中包含对回归测试原则（如“Given-When-Then”）的引用。

选择一个选项。

15. 你正在借助大语言模型，从原始数据中整理出具有可操作性的测试度量。这些度量包括测试进度，缺陷趋势和覆盖率，会以图表形式呈现，并配以文字说明。你的目标是优化测试过程，确保生成的度量准确、实用，且利益相关方能够轻松理解。

以下是用于指导 AI 的提示词初稿：

角色：作为一名测试经理。

上下文：你已获取来自测试工具的原始数据。

指令：根据原始数据生成测试进度度量，缺陷趋势度量和覆盖率度量。

输入数据：使用包含原始测试结果的附件。

约束：确保输出简明易懂。

输出格式：在仪表盘上显示度量。

现在要求你完善这个提示词。以下哪项改进最有助于提升大语言模型生成准确且实用的度量的能力？

A	明确该角色为专注于提供可行见解及决策支持的测试经理，以确保对测试数据进行全面分析。
B	添加一条指令，要求从生成的度量趋势中识别潜在风险，并给出风险影响评估及优先级。
C	扩充输出格式，增加一份以通俗易懂语言撰写的总结，对度量进行解读，并为利益相关者勾勒出后续步骤。
D	强调约束，即输出内容要便于利益相关方理解，整个回复需使用清晰直白的语言，避免技术术语。

选择一个选项。

16. 你的目标是为一个基于AI的系统创建测试用例，该系统存在测试结果参照物的问题，致使你无法确定预期结果。你只能依靠少数具有已知期望结果的现有测试用例。通过适当的分析，你已经确定了一组定义明确的转换规则，这些规则详细说明了输入的改变如何影响预期结果，并且这些规则适用于所有现有的测试用例。

你决定依赖生成式 AI，向特定的大语言模型提供以下信息：现有的测试用例及其输入和期望结果，转换规则的清晰描述，以及通过将这些规则精确应用于相关的现有测试用例来生成额外测试用例的指南。有了指定的信息，选定的大语言模型就能直接生成符合你预期的更多的测试用例。

在这种情况下，以下哪种提示技术最适合实现你的目标？

A	少样本提示。
B	提示词链。
C	元提示。
D	零样本提示。

选择一个选项。

17. 你正在借助生成式AI辅助测试一款娱乐软件应用程序。生成式AI模型针对用户交互场景生成测试用例，为应用程序编程接口（API）交互生成测试脚本，还会生成合成测试数据以处理边界用例。

为了有效地评估生成式 AI 模型的性能并优化提示词，以下哪组度和操作的组合最能确保全面的评估和改进？

A	评估测试用例的多样性，以确保存在多样的输入场景，并利用测试执行成功率来验证所生成的 API 测试脚本的功能。
B	运用准确性和完整性度量，依据娱乐软件的需求来验证测试用例，并借助时间效率，将AI生成的测试脚本与人工测试的工作量进行对比。
C	着重关注精确度，确保生成的测试数据符合娱乐软件的合规标准，同时通过上下文契合度和测试执行成功率，评估测试脚本的一致性和可用性。
D	对所有输出优先考量相关性和上下文契合度，使其与娱乐软件的需求保持一致，并纳入多样性度量，以扩大对边界用例的覆盖范围。

选择一个选项。

18. 以下哪一种评估及迭代完善提示词的技术，最适合用来查明为何大语言模型总是生成预期结果错误且与输入要求相悖的测试用例，进而为优化提示词、避免类似错误提供思路？

A	输出分析。
B	提示词的A/B测试。
C	调整提示词的长度与具体程度。
D	整合用户反馈。

选择一个选项。

19. 在大语言模型输出的语境中，“幻觉”指的是什么？

A	大语言模型未能准确执行多步推理过程而产生的逻辑错误。
B	因训练数据偏向某些观点而导致的大语言模型输出偏差。
C	大语言模型针对特定任务生成的不相关或与事实不符的输出内容。
D	大语言模型在测试生成任务中理解非英语视角内容时存在的局限性。

选择一个选项。

20. 你正在使用生成式AI为一款电子商务（网店）应用程序创建测试用例。项目简报中明确提到了以下功能：

- 购物车管理。
- 折扣码应用。
- 订单确认邮件生成。

基于上述细节，以下由 AI 生成的测试用例中，哪一个最有可能属于“幻觉”情况？

A	验证用户能够向购物车添加多个商品并进入结账环节。
B	验证用户在结账时无法使用已过期的折扣码。
C	验证用户成功下单后会收到确认邮件。
D	验证用户可以创建心愿单，以便日后保存心仪商品。

选择一个选项。

21. 在利用大语言模型执行测试任务时，使用清晰且结构化的输入数据格式，最直接相关的

好处是以下哪一项？

A	有助于减少针对测试任务对大语言模型进行微调的工作量。
B	有助于大语言模型针对测试任务生成歧义更少的输出。
C	有助于大语言模型针对测试任务生成与上下文关联更紧密的输出。
D	有助于大语言模型针对测试任务生成更具创意的输出。

选择一个选项。

22. 在推理过程中，通过缩小概率分布来降低大语言模型输出变化性的策略是哪一个？

A	提高学习率。
B	降低温度参数。
C	增大随机种子值。
D	减小随机种子值。

选择一个选项。

23. 以下关于在软件测试中使用生成式AI所涉及的数据隐私问题，哪项表述**不正确**？

A	生成式AI可能会在输出中无意间暴露敏感数据。
B	生成式AI工具可能在未获用户明确同意的情况下存储和处理敏感数据，进而导致数据滥用。
C	使用生成式AI工具若不遵守数据保护法规，如《通用数据保护条例》（GDPR），可能引发法律纠纷。
D	无论大语言模型基于何种数据训练，当其在生成合成测试数据时出现“幻觉”，都很可能泄露真实敏感数据。

选择一个选项。

24. 攻击者把伪造的测试结果，注入到一个用于推荐最佳测试覆盖策略的大语言模型的训练数据集中。这种描述**最符合**以下哪种攻击手段？

A	恶意代码生成。
B	数据泄露。
C	请求篡改。

D	数据投毒。
---	-------

选择一个选项。

25. 将以下每种针对大语言模型的攻击向量类型（1 - 4）与相应的示例（A - D）进行匹配：

1. 数据泄露。
 2. 请求篡改。
 3. 数据投毒。
 4. 恶意代码生成。
-
- A. 攻击者恶意篡改用于微调大语言模型数据集中的，与需求和测试用例之间可追溯性链接相关的数据，损害其依据需求生成测试用例的准确性。
 - B. 攻击者恶意编造并提供欺骗性提示词，诱导经微调后用于协助测试人员自动生成测试脚本的大语言模型，生成存在隐藏安全漏洞的易受攻击的测试脚本。
 - C. 攻击者恶意提供大量精心编造的提示词，诱导经微调后用于协助测试人员生成测试用例的大语言模型，使其无意间泄露从以往测试项目继承而来的机密 API 密钥。
 - D. 攻击者恶意向采用大语言模型进行视觉对比分析的视觉测试框架，提交精心修改的参考截图，诱使大语言模型在回归测试期间系统性地忽略真实的用户界面（UI）缺陷。

A	1C, 2D, 3A, 4B
B	1B, 2D, 3A, 4C
C	1D, 2C, 3B, 4A
D	1C, 2B, 3D, 4A

选择一个选项。

26. 以下哪种策略在基于生成式AI的软件测试场景中，能最有效地应对数据隐私风险？

A	使用多个大语言模型对测试结果进行评估与比较，以提升正确性。
B	将敏感测试数据替换为相应的匿名化版本。
C	允许对敏感测试数据进行无限制访问，以优化生成式AI模型的训练。

D	禁用敏感测试数据的加密功能，以使数据存储与传输流程更高效。
---	-------------------------------

选择一个选项。

27. 关于大语言模型使用对能源消耗和二氧化碳 (CO₂) 排放的影响，以下哪项描述是**正确**的？

A	图像生成任务消耗的能源比文本生成任务多得多，但产生的二氧化碳排放却更少。
B	由于算法得到优化，由生成式AI驱动搜索比传统网络搜索消耗的能源要少得多。
C	图像生成任务因计算复杂度更高，所以消耗的能源比文本生成任务多得多。
D	文本生成任务消耗的能源极少，即便数百万用户同时使用，也不会造成明显的能源消耗。

选择一个选项。

28. 在软件测试中使用生成式AI时，以下哪两项标准（或其部分内容）**最为相关**？

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

选择二个选项。

29. 在基于大语言模型的测试应用里，以下哪个组件负责把用户输入和结构化且语义相近的数据整合起来，以便为大语言模型准备提示词？

A	后端。
B	前端。
C	认证组件。
D	后处理组件。

选择一个选项。

30. 你是一名负责测试银行应用程序的测试人员，该应用程序具备用户登录、账户管理以及安全交易等功能。系统文档（包括 API 规范和安全要求）存储在向量数据库中，而历史测试用例则存储于关系型数据库中。你的任务是运用检索增强生成（RAG）框架生成测试用例，确保其与最新的规范和要求相符。

在此场景下，以下哪个选项**最恰当地**体现了对 RAG 框架的运用？

A	提交一个指定单个待测试功能的查询。RAG 框架会从向量数据库中检索相关规范与要求，将其与历史测试用例相结合，通过大语言模型自动生成精确且具有上下文感知的测试用例。
B	提交一个指定所有待测试功能的查询。RAG 框架会从向量数据库中检索相关规范与要求，将其与历史测试用例相结合，通过大语言模型自动生成精确且具有上下文感知的测试用例。
C	利用 RAG 框架从关系型数据库检索历史测试用例，并从向量数据库检索安全要求。手动审查检索到的信息后，优化针对大语言模型的查询，以生成目标明确的测试用例。
D	依靠大语言模型的内部训练数据生成测试用例，同时使用 RAG 框架进行参考检索，但不将检索到的信息直接整合到生成过程中。

选择一个选项。

31. 以下哪个选项**最恰当地**描述了由大语言模型驱动的自主及半自主智能体，为自动化测试过程带来的改进？

A	通过均衡运用单智能体和多智能体系统，它们能提高自动化测试过程的效率与质量。
B	虽然会降低效率，但它们能通过增加复杂的验证检查，提高自动化测试过程的质量。
C	凭借在不同程度人机交互条件下运行的能力，它们能提高自动化测试过程的效率与质量。
D	它们能提高自动化测试过程的效率与质量，同时省去这些测试过程中的验证步骤。

选择一个选项。

32. 以下关于针对特定测试任务对语言模型进行微调的表述，哪一项是**错误的**？

A	微调指的是在特定任务数据上对预训练模型进行训练，以提升其性能和领域知识。
B	微调通过用特定任务推理取代语言模型的通用知识，为其赋予新能力，同时确保不会发生过拟合。

C	微调通过使用目标数据集修改预训练模型的参数，以使其适配特定领域或任务。
D	微调需要高质量、面向特定任务的数据集，以避免出现偏差或不准确的结果。

选择一个选项。

33. 在为测试任务部署和管理大语言模型时，以下哪一项最恰当地描述了大语言模型运维（LLMOps）的主要关注点？

A	避免在测试过程中对生成式AI产生依赖。
B	在大语言模型的整个生存周期内，从隐私、安全及成本等方面考量，对其进行有效管理。
C	将大语言模型的使用限制在基于聊天机器人的测试解决方案中，从而降低复杂性。
D	实现所有测试任务的完全自动化，而无需人工监督。

选择一个选项。

34. 以下关于影子AI的表述，哪一项是**正确的**？

A	影子AI会强制执行组织数据政策和通用AI法规的依从性要求。
B	影子AI使得AI工具中不再需要明确的许可协议。
C	影子AI降低了知识产权纠纷的风险。
D	影子AI可能导致对敏感信息的未授权访问。

选择一个选项。

35. 在为软件测试定义生成式AI策略时，需要考虑的一个关键要素是什么？

A	准备培训计划，以确保团队成员获得其所使用的每个大语言模型的特定认证。
B	选择能够与现有测试环境和测试工具进行恰当集成的大语言模型。
C	确保提供尽可能多的输入数据，以提高获得有效的大语言模型输出的可能性。
D	收集标准的监督式机器学习度量，以评估大语言模型输出的有效性。

选择一个选项。

36. 当测试组织为特定测试任务选择合适的大语言模型时，以下哪项陈述**最恰当地**描述了其中一项关键标准？

A	关键选择标准包括：在代码生成中，参照大语言模型在代码生成方面公开可用的基准，评估该大语言模型针对测试任务的性能。
B	关键选择标准包括：评估诸如运行大语言模型所需计算资源相关的周期性成本。
C	关键选择标准包括：参照公开可用的社区基准，评估该大语言模型，以确保与之完全兼容。
D	关键选择标准包括：评估周期性成本，例如与概念验证相关的成本，该验证旨在证明大语言模型适用于测试任务。

选择一个选项。

37. 测试组织采用生成式 AI 的关键阶段有哪些？

A	探索、启动与使用定义、应用与迭代。
B	认知、使用优先级设定、性能监测。
C	规划、试验、评估与优化。
D	培训、测试、实施与推广。

选择一个选项。

38. 以下哪项**最能体现**测试人员在测试过程中有效使用大语言模型所需的知识和/或技能？

A	掌握专门技术，旨在防止大语言模型在执行特定测试任务时产生幻觉或推理错误。
B	选择并实施合适的测试自动化方法（如关键词驱动的测试自动化）以实现测试过程自动化。
C	根据大语言模型是否具备适应或定制能力以执行特定测试任务等标准，选择最合适的模型。
D	确保用于大语言模型开发的验证数据和测试数据具有最高质量。

选择一个选项。

39. 为了有力支持生成式 AI 的应用，培养测试团队技能的**最佳方式**是什么？

A	主要依托外部专家课程并结合实践操作，目标是一次性将人工智能融入所有日常测试任务中。
B	鼓励在无结构化过程的情况下，自主对各类大语言模型进行试验。
C	采用实践性、渐进式的学习过程，借助指导性练习、同伴互学以及知识共享社区来推进。

D	主要依托外部专家的理论课程，旨在根据实际学习情况逐步将人工智能整合到日常测试任务中。
---	--

选择一个选项

40. 以下哪项**最恰当地**描述了在测试组织内，采用生成式 AI 进行软件测试对测试人员和测试经理的角色与职责的影响？

A	测试人员的工作重心从手动设计测试用例，转变为引导并验证由AI生成的测试件。
B	测试经理的工作重心从管理测试项目，转变为了解生成式AI技术的内在运行机制。
C	测试人员的工作重心从手动设计测试用例，转变为监督基于AI的测试过程。
D	测试经理的工作重心从依靠人力，转变为完全依靠生成式AI来提高测试任务的效率。

选择一个选项。

中国软件测试认证委员会

附录：附加题

A1. 考虑以下实施检索增强生成（RAG）所涉及的步骤（在左侧）：

步骤	答案
清理并处理文档块。	首先
使用检索到的文档块和大语言模型生成响应。	
将大文档拆分为较小的文档块。	
将嵌入内容存储于向量数据库中。	
基于语义相似度检索相关文档块。	最后

将这些步骤按先后顺序从第一步到最后一步进行排序。

A2. 将每种提示技术（在左侧）与其在软件测试中的合适应用示例（在右侧）进行匹配：

提示词链。	将测试分析任务分解为较小步骤，这些步骤需要与大语言模型进行迭代交互并由人工验证。
少样本提示。	利用大语言模型，依据预先设定的示例，从用户故事生成 Gherkin 风格的测试用例。
元提示。	在无示例的情况下，借助大语言模型的内在推理能力，根据风险对测试用例进行优先级排序。
零样本提示。	与大语言模型交互，迭代优化提示，以便从模糊需求中创建测试结果参照物。

请将每种提示技术与它在软件测试中的恰当应用示例进行配对。每个条目均需完成配对，且不得与多个条目配对。

A3. 请考虑在测试组织中采用生成式 AI 时涉及的以下活动（左侧）和关键阶段（右侧）：

活动	关键阶段
识别实际用例并确定其优先级。	探索。
对测试人员进行生成式 AI 培训。	启动与使用定义。

为测试人员提供大语言模型的使用权限。		应用与迭代。
通过试验初始用例，让测试人员熟悉生成式 AI。		
在生成式 AI 全面整合后，管理测试过程的演进。		

将每个活动（条目）分配至关键阶段（组别）。任何组别不得留空，且每个条目不得分配至多个组别。