

认证测试工程师 专业领域人工智能测试 模拟卷 A-答案 (大纲 1.0 版)

版本号: EN1.0_CN1.0

发布日期: 2024 年 10 月 18 日

国际软件测试认证委员会

ISTQB

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。

版权申明

版权声明©国际软件测试资格委员会(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。

本文作者已将版权转让给 ISTQB®。本文作者(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:

非商业用途,如果来源确认,可复制本文内容。

如果本文作者和 ISTQB®被确认为此份考试样题的来源和版权所有人,则任何经认证的培训机构可在其培训课程中使用此份考试样题,且此类培训课程的任何广告只允许在从 ISTQB®认可的成员委员会收到培训材料的官方认证后才能发布。

如果本文作者和 ISTQB®被确认为此份考试样题的来源和版权所有人,则任何个人或团体都可以在文章或书籍中使用此份考试样题。

在没有事先获得 ISTQB®书面批准的情况下,禁止任何形式使用此份考试样题。

任何 ISTQB®认可的成员委员会都可以翻译本份考试样题,前提是将上述版权声明复制到考试样题的翻译版本中。

文档责任

ISTQB®检查工作组负责本文件。

致谢

本文件是由 ISTQB® 的核心团队制作，成员包括:Klaudia Dussa-Zieger, Werner Henschelchen, Vipul Koch, Qin Liu, Stuart Reid, Kyle Siemens 和 Adam Leon Smith。

核心团队感谢考试工作组审查小组、教学大纲工作组和国家委员会的建议和投入。

本文件由 ISTQB® 的核心团队维护，该团队由教学大纲工作组和考试工作组组成。

人工智能测试文档中文版本 1.0 翻译参与者（按姓氏拼音排序）：

曹栋、陈嘉诚、陈希、陈智迪、高方原、高蕊、冷炜、刘惠、刘佳钰、任亮、商超博、张希婷

人工智能测试文档中文版本 1.0 QA 评审参与者（按姓氏拼音排序）：

陈晟、丁国富、董昕、范鸿飞、梁静、吴洁

致谢企业：

上海均瑜管理咨询有限公司



均瑜咨询
JY CONSULTING

修订历史

版本	日期	备注
1.0	2021/10/01	GA 版本
EN1.0-CN1.0	2024/10/18	英文版 1.0 本地化完成

目录

版权申明.....	2
文档责任.....	2
致谢.....	3
修订历史.....	4
目录.....	5
引言.....	6
文档目的.....	错误!未定义书签。
说明	6
参考答案.....	7
答案解析.....	8
附录：附加问题的答案.....	24

引言

文档目的

本样题集的题目、答案和相关论证是由相关领域专家和经验丰富的试题作者团队创建，旨在协助ISTQB®成员委员会和考试委员会进行试题编制工作。

这些试题不能直接在任何正式考试中使用，但应作为试题编写作者的指导。鉴于形式和主题的多样性，这些样本试题应该为各成员委员会提供更多的想法，以便了解如何为自己组织的考试创建好的题目和适当的答案集。

说明

在本文档中，你可以找到：

- 答案表，包括每个正确答案：
 - k 级、学习目标和分值
- 答案集，包括所有问题：
 - 正确答案
 - 每个回答(答案)选项的详细解析
 - k 级、学习目标和分值
- 额外的答案集，包括所有的试题[不适用于所有样题]：
 - 正确答案
 - 每个回答(答案)选项的详细解析
 - k 级、学习目标和分值
- 问题包含在另一份文件中

参考答案

试题序号	正确答案	学习目标	试题等级	分数
1	c	AI-1.1.1	K2	1
2	c	AI-1.4.1	K1	1
3	b	AI-1.6.1	K2	1
4	b	AI-1.8.1	K2	1
5	a	AI-2.2.1	K2	1
6	d	AI-2.4.1	K2	1
7	b	AI-2.6.1	K2	1
8	a	AI-2.8.1	K1	1
9	b	AI-3.1.1	K2	1
10	a	AI-3.1.3	K2	1
11	c	AI-3.3.1	K3	2
12	c	AI-3.5.1	K2	1
13	c	AI-4.1.1	K2	1
14	d	AI-4.3.1	K2	1
15	b	AI-4.4.1	K2	1
16	b	AI-4.5.1	K1	1
17	c	AI-5.1.1	K3	2
18	a	AI-5.2.1	K2	1
19	b	AI-5.4.1	K4	2
20	b	AI-6.1.1	K2	1

试题序号	正确答案	学习目标	试题等级	分数
21	b	AI-6.2.1	K2	1
22	d	AI-7.1.1	K2	1
23	a	AI-7.3.1	K1	1
24	d	AI-7.4.1	K2	1
25	b, e	AI-7.7.1	K4	2
26	c	AI-8.1.1	K2	1
27	a	AI-8.3.1	K2	1
28	b	AI-8.5.1	K2	1
29	b, c	AI-8.8.1	K4	2
30	b	AI-9.1.1	K2	1
31	c	AI-9.2.1	K2	1
32	a	AI-9.3.1	K2	1
33	b	AI-9.5.1	K3	2
34	a	AI-9.6.1	K2	1
35	b, c	AI-9.7.1	K4	2
36	a	AI-10.1.1	K2	1
37	b	AI-11.2.1	K2	1
38	c	AI-11.3.1	K2	1
39	b	AI-11.4.1	K2	1
40	b	AI-11.5.1	K2	1

答案解析

试题序号	正确答案	解释理由	学习目标	试题等级	分数
1	c	<p>a) 不正确。许多职业的人可能会因为基于人工智能的系统而失去工作，但这只是进步，而不是“人工智能效应”</p> <p>b) 不正确。对于一些电脑游戏来说，基于人工智能的系统可以胜过人类，但几乎没有证据表明这类游戏的受欢迎程度有所下降。</p> <p>c) 正确。“人工智能效应”是指随着技术的进步，人工智能的定义发生了变化。基于规则的医疗诊断系统是 20 世纪 70 年代和 80 年代人工智能流行的应用实例，但如今已经普遍不被认为是人工智能。</p> <p>d) 不正确。电影观众轻信机器人杀手将会统治世界，并不是“人工智能效应”。</p>	AI-1.1.1	K2	1
2	c	<p>a) 不正确。支持向量机是机器学习的一种形式。</p> <p>b) 不正确。决策树是机器学习的一种形式。</p> <p>c) 正确。目前没有像进化推理这样的人工智能技术。这一术语在几十年前，人工智能被认为是基于生物进化规则的讨论中偶尔会被用到。</p> <p>d) 不正确。贝叶斯优化是机器学习的一种形式。</p>	AI-1.4.1	K1	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
3	b	<p>a) 不正确。训练机器学习模型和从该模型推理两种活动是完全不同的，所以通常没有理由在相同的处理器上执行它们。</p> <p>b) 正确。图形处理器是为了使用数千个内核并行处理图像而设计的，这接近于一个基于人工智能的计算机视觉系统所需要的功能，该系统最有可能以神经网络的形式实现。</p> <p>c) 不正确。在图形处理器支持有限的台式机上训练、评估和测试一个简单的深度学习系统仍然是可能的——所以不需要特定的人工智能芯片，但如果有的话速度会快得多。</p> <p>d) 不正确。许多基于人工智能的系统并不关注精确的计算，而是关注概率的确定，因此具有许多位的处理器的精度通常是不必要的。</p>	AI-1.6.1	K2	1
4	b	<p>a) 不正确。这个问题提到了预训练模型的质量良好，因此这种风险应该可以忽略不计。</p> <p>b) 正确。用于训练模型的数据应该与用于进行预测的数据相似。</p> <p>c) 不正确。在这种情况下，性能通常不是问题。</p> <p>d) 不正确。在这种情况下，可解释性一般不是问题，也不能通过查看训练数据及其格式来实现。</p>	AI-1.8.1	K2	1
5	a	<p>a) 正确。该需求定义了人工干预，定义了系统自动工作结束则开展人工干预。</p> <p>b) 不正确。这一要求规定了系统如何进行自我学习所需的功能。</p> <p>c) 不正确。该要求规定系统将如何管理概念漂移，在这种情况下，最有可能是由房地产市场变化引起的。</p> <p>d) 不正确。这是指定一个适应性要求——对系统进行更改所需的最长时间。</p>	AI-2.2.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
6	d	<p>a) 不正确。如果用户故意破坏基于人工智能的系统的自我学习，可能会导致偏差。</p> <p>b) 不正确。当训练数据不能正确地匹配系统将要应用的对象时，可能会造成偏差。例如，员工通常会比退休者年轻。</p> <p>c) 不正确。当训练数据不能正确地匹配系统将要应用的对象时，可能会造成偏差。例如，大多数使用信用卡的人已经被认为是信誉良好的，这是一个典型的样本偏差的例子。</p> <p>d) 正确。如果算法不能被解释，那么它就缺乏可解释性，但这并不意味着它是有偏见的或无偏见的。</p>	AI-2.4.1	K2	1
7	b	<p>a) 不正确。似乎该工具在没有有害影响的情况下实现了它的两个目标，所以这不太可能是“奖励黑客”。</p> <p>b) 正确。如果系统实现了一个目标而同时损害了其他目标，这可能是“奖励黑客”，在这种情况下，受损者需要唤醒。</p> <p>c) 不正确。奖励黑客并不是支付给人工智能开发者的一种形式。</p> <p>d) 不正确。有些基于人工智能的游戏系统是由奖励功能驱动的，但这并不是所谓的“奖励黑客”。</p>	AI-2.6.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
8	a	<p>考虑到给定的属性：</p> <ul style="list-style-type: none"> · 概率性——对于安全相关系统来说，这是一个明确的问题，因为这会导致不确定性 · 可解释——安全相关系统通常需要 · 不公平——不理想，但有时不可避免——对于与安全相关的系统来说不是一个特殊的问题 · 不确定性——安全相关系统的一个明确问题 · 定性-安全相关系统通常需要 <p>因此，I 和 IV 是安全相关系统中最有问题的属性。</p> <p>a) 正确</p> <p>b) 不正确</p> <p>c) 不正确</p> <p>d) 不正确</p>	AI-2.8.1	K1	1
9	b	<p>a) 不正确。在有监督学习环境下的回归验证，通常是在机器学习模型输出数据结果时进行。</p> <p>b) 正确。分类是指将机器学习模型的输入数据分类为几个预定义的类之一。</p> <p>c) 不正确。在有监督学习中，训练数据需要贴上标签，但这项活动不称为分类，只是贴标签。</p> <p>d) 不正确。回归验证是当机器学习模型的输出是数值的，但输出不是多个类的时候进行。</p>	AI-3.1.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
10	a	<p>a) 正确。花费的金额可视为该系统的奖励功能，系统会改变其行为以增加花费的金额。</p> <p>b) 不正确。该应用程序使用的文本可以被视为是源语言及其的“正确”翻译。因此，它依赖于一种有监督学习的形式，没有提到奖励功能。</p> <p>c) 不正确。该系统将人的质量控制作为“黄金”标准的一种形式，因此其依赖于一种有监督学习的形式。</p> <p>d) 不正确。没有建议使用任何奖励功能，相反，最有可能的是预测系统根据过去的经验来确定缺陷。因此，它可能也依赖于监督式学习系统。</p>	AI-3.1.3	K2	1
11	c	<p>a) 不正确。无监督学习系统应该能够识别出拥堵的区域，但这些并不能提供解决方案。</p> <p>b) 不正确。回归验证解决方案不太可能为我们提供我们想要的结果，因为预测的单个行程的速度不会为整个城市的拥堵提供整体解决方案。</p> <p>c) 正确。一个持续改进的强化学习系统，其奖励功能是基于较低的拥堵水平，作为成功的衡量标准，对于这种类型的系统是有效的。</p> <p>d) 不正确。这个解决方案取决于志愿者提交的主观意见，这些意见很可能会导致解决方案随着系统采用最喜欢的路线而来回变化，而变得拥挤。</p>	AI-3.3.1	K3	2
12	c	<p>a) 不正确。该模型在验证数据上表现良好，因此不会出现拟合不足的情况。</p> <p>b) 不正确。概念漂移是指经过模型训练和验证阶段后的变化。</p> <p>c) 正确。试验数据的表现不佳，验证数据的表现较好，表明存在拟合过度。</p> <p>d) 不正确。较差的验收标准应该与不同的数据集一致，因此不太可能导致验证数据的测试结果与独立测试数据之间的差异。</p>	AI-3.5.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
13	c	a) 不正确。数据匿名化操作不需要了解机器学习算法。 b) 不正确。非结构化数据不是一个挑战。图像、音频、自由文本都是非结构化数据的示例。 c) 正确。 多达 36%的机器学习工作流的工作可能花费在数据准备上。 d) 不正确。可扩展性通常是部署时的需求，而不是训练时的需求。	AI-4. 1. 1	K2	1
14	d	a) 不正确。由于基于某些学习算法的模型可以用该数据进行训练，但不能用特定一种算法进行训练，这表明数据是正确的。 b) 不正确。由于基于某些学习算法的模型可以用数据进行训练，但不能用特定一种算法进行训练，这表明没有丢失数据。 c) 不正确。由于基于某些学习算法的模型可以用数据进行训练，但不能用特定一种算法进行训练，这表明数据被正确标注。 d) 正确。因为基于某些学习算法的模型可以用该数据进行训练。然而，如果它对特定一种算法不起作用，很有可能是数据的数量不足以满足该特定算法。	AI-4. 3. 1	K2	1
15	b	a) 不正确。数据隐私和安全问题没有得到处理。因此，该产品不会避免安全问题。 b) 正确。错误标注的数据会降低机器学习模型的准确性。 c) 不正确。模型不适合的原因是数据不正确或不公平，而不是标注错误的数据。 d) 不正确。有偏差模型产生于不完整的数据，不平衡的数据，不公平的数据，缺乏多样性的数据，或重复的数据，而不是标注错误的数据。	AI-4. 4. 1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
16	b	a) 不正确。众包是指使用大量人员提供一些工作。在这种情况下，只有一个人执行任务。 b) 正确。这里增强是通过转换现有的带标注的数据来实现的。 c) 不正确。人工智能并没有被用于标注数据。 d) 不正确。机器学习的工程师没有把这个任务外包给第三方。	AI-4.5.1	K1	1
17	c	a) 不正确。正确的公式和计算方法见选项 c。 b) 不正确。正确的公式和计算方法见选项 c。 c) 正确。精确度公式= 真阳性/ (真阳性+假阳性) *100 = 78/(78+22) = 78/100 *100 d) 不正确。正确的公式和计算方法见选项 c。	AI-5.1.1	K3	2
18	a	a) 正确。这是一个监督回归问题，因为模型输出一个连续的值，即将产生的电量使用 R-判定系数或均方误差/均方根误差度量。 b) 不正确。这是一种分类度量。 c) 不正确。这是一种分类度量。 d) 不正确。这是一种分类度量。	AI-5.2.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
19	d	<p>a) 不正确。在这种情况下，当预期的类别不平衡且不饥饿类占主导地位时，准确性就没有用了。</p> <p>b) 正确。因为假阳性(给狗过度喂食)的成本很高(严重的健康问题)，所以需要使用精确度。</p> <p>c) 不正确。当阳性不应该被遗漏时召回率是有用的。在这种情况下，精确性也很重要(见 b)，因此仅召回率不是很有用。F1 分数是更好的选择。</p> <p>d) 不正确。当预期的类别中存在不平衡时并且精确度和召回率同样重要时，F1 分数是有用的，但在这种情况下，精确度似乎比召回率更重要。</p>	AI-5.4.1	K4	2
20	b	<p>a) 不正确。神经网络没有层次结构。</p> <p>b) 正确。与人类大脑一样，人工神经网络是由相互连接的神经元组成的。为了计算激活值，每个神经元被分配一个偏置，每个连接被分配一个权值。</p> <p>c) 不正确。神经网络由几层组成，误差通过网络向后传播，但神经网络的每一层只连接到下一层(而不是相互连接)。</p> <p>d) 不正确。神经网络是由多层神经元组成的，但激活值是基于前一层(不是同一层)的神经元。</p>	AI-6.1.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
21	b	<p>a) 不正确。值变更覆盖率是一种测量被激活的神经元的比例，它们的激活值的差异超过一个预设的变化量。它与神经网络的总体输出无关。</p> <p>b) 正确。阈值覆盖率在测试中激活的神经元比例的值大于预设阈值。</p> <p>c) 不正确。每次神经网络“运行”时，所有的神经元都被潜在地“激活”，然而神经元输出的值发生了变化，这是由神经元覆盖率(通过大于零的值实现覆盖)来衡量的。</p> <p>d) 不正确。符号变更覆盖率是一种同时激活阳性和阴性激活值，但不是零激活值的神经元比例的度量。</p>	AI-6.2.1	K2	1
22	d	<p>a) 不正确。这是测试结果参照物的一个特定需求，所以通常不会引起测试挑战。</p> <p>b) 不正确。这可能是一个很难实现的需求，但不应该产生测试挑战。</p> <p>c) 不正确。这是一个可测试的需求。</p> <p>d) 正确。如果不定义所有的人类情感以及系统如何模仿它们，这个需求的测试是极其复杂的。</p>	AI-7.1.1	K2	1
23	a	<p>a) 正确。为使用大量高速数据的人工智能系统寻找数据可能很困难。</p> <p>b) 不正确。从多个来源获取一致的数据可能很困难。</p> <p>c) 不正确。单独查找数据是一种很好的做法，因为它可以避免数据科学家遇到常见的故障。</p> <p>d) 不正确。从公共网站获取数据很简单。</p>	AI-7.3.1	K1	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
24	d	<p>a) 不正确。决策的速度与准确性无关。</p> <p>b) 不正确。人类做出的伦理选择与测试人工智能系统无关。</p> <p>c) 不正确。人类决策的准确性与被测试的人工智能系统相关的，因为系统可能会提出人类批准或审查的建议。</p> <p>d) 正确。由基于人工智能系统的建议支持的人类决策的质量可能低于没有系统建议的人类决策，这应该在测试中考虑。</p>	AI-7.4.1	K2	1
25	b, e	<p>a) 不正确。概念漂移在部署后进行测试。</p> <p>b) 正确。对抗测试很重要，因为需求声明不应该针对漏洞对系统进行测试。</p> <p>c) 不正确。可扩展性测试并没有作为需求之一被提及。它们是相互连接的独立系统，与任何其他系统无关。</p> <p>d) 不正确。公平性是在训练中使用有正面偏见的数据。既然这里没有正面歧视的案例，公平测试就无关紧要了。</p> <p>e) 正确。数据管道测试是必需的，因为图像可以有各种格式和分辨率。对于要训练的模型，所有图像都应该有相同的格式，因此这个测试是重要的。</p>	AI-7.7.1	K4	2
26	c	<p>a) 不正确。一个需要定期再训练的系统不能被描述为自学。</p> <p>b) 不正确。需要定期发布的系统不能描述为自学，因为它需要频繁发布以适应变化。</p> <p>c) 正确。对系统本身进行更改的测试可能开始会失败，即便它们之前通过了。</p> <p>d) 不正确。一个需要人工操作的系统不太可能是自我学习的。</p>	AI-8.1.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
27	a	<p>a) 正确。存在偏差的用户不需要测试系统是否存在偏差，因为他们不能帮助确定系统的行为是否存在偏差。</p> <p>b) 不正确。当测试偏差时，测量测试输入如何改变测试输出是很重要的，因为它可以显示系统如何偏向或反对特定的输入。</p> <p>c) 不正确。在测试偏差时，测量生产输入如何改变生产输出是很重要的，因为在生产中可能会看到不同的结果。</p> <p>d) 不正确。如果偏差是基于“隐藏”变量，那么在测试偏差时，获取外部数据源是至关重要的。</p>	AI-8.3.1	K2	1
28	b	<p>a) 不正确。偏差通常与系统的复杂性无关。</p> <p>b) 正确。理解系统如何工作并创建足够的测试以实现有效的覆盖是基于人工智能系统的复杂性带来的挑战。</p> <p>c) 不正确。伦理通常与基于人工智能的系统复杂性无关。</p> <p>d) 不正确。很难找到有代表性的数据来训练模型与测试或基于人工智能的系统复杂性无关。</p>	AI-8.5.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
29	b、c	<p>考虑每个属性：</p> <p>a) 不正确。适应性 - 是系统被修改的能力(通常是为了继续满足功能性和非功能性需求)。我们没有理由相信系统的操作环境会发生很大的变化，因此也没有理由相信系统必须改变。</p> <p>b) 正确。偏差 - 用于培训的数据偏向于妇女(25,000 对 5,000)和特定年龄组，因此需要仔细考虑偏差。</p> <p>c) 正确。可解释性 - 研究结果可能会在医学上和经济上影响已确定的弱势患者。他们应该能够看到为什么他们被贴上易受影响的标签，这样他们就可以确保他们被正确选择，并作为与数据隐私相关的解释要求的一部分。</p> <p>d) 不正确。灵活性 - 是系统改变其行为的能力，但在这一点上没有理由认为这个系统将需要在初始规范之外使用，因此它不期望有一个需要改变其行为。</p> <p>e) 不正确。自主性 - 是系统在没有人为干预的情况下持续工作的能力。没有理由认为该体系将不得不在不进行干预的情况下持续工作。</p> <p>因此，偏差和可解释性是给定场景的两个最重要的属性。</p> <p>a) 不正确。</p> <p>b) 正确。</p> <p>c) 正确。</p> <p>d) 不正确。</p> <p>e) 不正确。</p>	AI-8.8.1	K4	2

试题序号	正确答案	解释理由	学习目标	试题等级	分数
30	b	a) 不正确。这不是验证，因为可开发的输入正在被发现，并且正在使用这些输入进行再训练。 b) 正确。这是对抗测试的一个例子。 c) 不正确。在这种情况下没有测试任何数据管道。 d) 不正确。本例中没有执行可扩展性测试。	AI-9.1.1	K2	1
31	c	a) 不正确。A/B 测试并不适用于组合测试。 b) 不正确。所有的组合在实践中几乎都是不可能的，这是由接近无限数量的潜在组合造成的。 c) 正确。结对测试最适合于减少组合的数量，而不会牺牲太多的缺陷检测。 d) 不正确。背靠背测试对于组合测试没有用处。	AI-9.2.1	K2	1
32	a	a) 正确。这是背靠背测试的一个例子，其中非人工智能系统被用作伪测试结果参照物。 b) 不正确。在 A/B 测试中，我们使用被测系统的一个变体与被测系统进行比较。 c) 不正确。伪测试结果参照物和被测系统的资源和非功能性特征可能是不同的，因此备用系统不能用于非功能性测试。 d) 不正确。这是一个背靠背测试的例子，而且伪测试结果参照物和被测系统的资源和非功能特征可能会有所不同。	AI-9.3.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
33	b	<p>后续测试用例 T1 与原测试用例的区别在于 3D 摄像机需求的变化;现在它更具体了。必须包括一个 3D 摄像机。这意味着后续的期望结果最多只能包括原来的测试结果(之前推荐的带有 3D 摄像头的手机)。</p> <p>后续测试用例 T2 与源测试用例的不同之处在于 3D 摄像机需求的变化;它也更具体。不应包含 3D 摄像机。这意味着后续期望结果最多只能包括原来的测试结果(之前推荐的没有 3D 摄像头的手机)。</p> <p>由于 T1 列出了带有 3D 摄像头的手机,源测试用例中剩下的手机必须是那些没有 3D 摄像头的手机——因此它们应该在 T2 中。</p> <p>因此, T1 和 T2 的组合应该包含源测试用例中的所有摄像机,但两者之间不存在重叠。</p> <p>a) 不正确。</p> <p>b) 正确。</p> <p>c) 不正确。</p> <p>d) 不正确。</p>	AI-9.5.1	K3	2
34	a	<p>a) 正确。这是探索性数据分析,它是一种探索性方法。</p> <p>b) 不正确。这是脚本测试。</p> <p>c) 不正确。这是基于检查表的测试。</p> <p>d) 不正确。计算机机器学习功能表现度量不是探索性测试。</p>	AI-9.6.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
35	b, c	<p>a) 不正确。A/B 测试在比较两个变体以确定新变体是否优于旧变体时是最有用的。</p> <p>b) 正确。背靠背测试使用类似的产品作为伪测试结果参照物进行测试。</p> <p>c) 正确。对抗测试在这里很重要，因为它被用于非常重要的目的，而对抗性数据可能会造成伤害。</p> <p>d) 不正确。虽然状态转换测试可能有用，但场景中没有任何提示；因此，它不是最相关的技术。</p> <p>e) 不正确。这种测试适用于模型测试阶段的分类问题。非分类问题不适合在系统测试阶段进行。</p>	AI-9.7.1	K4	2
36	a	<p>a) 正确。可能需要为人工智能环境提供可解释性机制。</p> <p>b) 不正确。传统系统通常需要模拟器和虚拟环境。</p> <p>c) 不正确。传统系统也可能需要大量的数据。</p> <p>d) 不正确。许多其他系统也可能需要图形处理器，例如游戏。</p>	AI-10.1.1	K2	1
37	b	<p>a) 不正确。如果有少量缺陷需要分类，而且没有历史数据，那么人工智能就没有训练数据可用。遗传算法和神经网络可以用于测试生成，甚至可以结合使用。聚类分析生成适用于测试生成的结果。</p> <p>b) 正确。当一个小型应用程序报告了大量的缺陷时，最有可能识别重复的优点和机会。</p> <p>c) 不正确。当缺陷报告中提供的数据最少时，工具的有效性就会降低，因为算法可以获得的数据更少。</p> <p>d) 不正确。人工智能推荐开发者修复缺陷需要基于历史数据。然而由于一个新的开发团队正在接管，在获得历史数据之前任何建议都是不准确的。</p>	AI-11.2.1	K2	1

试题序号	正确答案	解释理由	学习目标	试题等级	分数
38	c	<p>a) 不正确。测试章程提供了探索性测试的焦点，它很少生成测试用例，即使基于人工智能的工具可以解释它。</p> <p>b) 不正确。流程图可以用来生成测试，但它需要机器可读，而不是简单的图片。</p> <p>c) 正确。网络服务器日志可以反映系统的生产使用情况，并为人工智能生成测试提供一种方法。</p> <p>d) 不正确。不太可能使用崩溃报告，因为它们将描述意外的失败，而不是应用程序执行的功能。</p>	AI- 11.3.1	K2	1
39	b	<p>a) 不正确。回归测试优化的目标是减少测试集的大小、优先级排序或增加测试集，而不是减少误检。</p> <p>b) 正确。回归测试套件的优化是通过分析关于先前测试执行的信息来执行的。</p> <p>c) 不正确。根据大纲的 11.4 节，回归测试优化通常使用以前的测试执行数据执行。使用遗传算法来创建新的测试不太可能实现优化回归测试套件的目标。</p> <p>d) 不正确。同时考虑回归测试和概念漂移是很重要的，但是根据 11.4 节，概念漂移与使用人工智能的回归测试优化无关。</p>	AI- 11.4.1	K2	1
40	b	<p>a) 不正确。虽然自然语言处理是一种人工智能应用程序，但它并不用于缺陷预测。</p> <p>b) 正确。缺陷预测是通过在相同或相似的代码基础上寻找代码/过程/人员度量与缺陷之间的相关性来执行的。</p> <p>c) 不正确。缺陷预测的目标不是识别具有假阳性结果的缺陷。分析它们没有什么价值。</p> <p>d) 不正确。缺陷预测不涉及使用规则扫描代码。这是静态分析。</p>	AI- 11.5.1	K2	1

附录：附加问题的答案

问题编号 (#)	正确答案	解释/原因	学习目标 LO	K	分数
1	a	<p>a) 正确。通用人工智能表现出与人类相当的智能行为，而一个自主机器人拥有足够广泛的技能，可以在完全不同的环境中担任工作人员，这很可能就是普通人工智能。</p> <p>b) 不正确。“奇点”是指人工智能超越人类的时刻，而不是它们处于相似水平的时候。</p> <p>c) 不正确。执行测试管理将被大多数人认为是一个单一的专门的任务集，并将被归类为狭义的人工智能。</p> <p>d) 不正确。狭义人工智能能够执行单一的专门任务，人工智能是否能接入互联网无关紧要。</p>	AI-1.2.1	K2	1
2	b	<p>a) 不正确。该系统似乎是描述客户的聚类，将执行无监督学习系统；因此，它很可能是基于人工智能的。</p> <p>b) 正确。没有证据表明这个系统实现了人工智能。</p> <p>c) 不正确。这个系统似乎通过听录音来编程；因此，它很可能是基于人工智能的。</p> <p>d) 不正确。这个系统似乎描述了一种基于 x 射线异常的有监督学习形式；因此，它很可能是基于人工智能的。</p>	AI-1.3.1	K2	1
3	d	<p>a) 不正确。这是一个用于 Python 编程语言的免费机器学习软件库。</p> <p>b) 不正确。这是微软认知工具包 (CNTK)，一个开源的深度学习工具包。</p> <p>c) 不正确。这是亚马逊用于 AWS 的深度学习开源框架。</p> <p>d) 正确。目前还没有这个名字的人工智能开发框架。</p>	AI-1.5.1	K1	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
4	c	<p>a) 不正确。图像分类器似乎是为一个高度特定的问题领域而构建的，不太可能作为人工智能即服务被广泛使用。</p> <p>b) 不正确。虽然避障系统是由第三方提供的，但它很有可能是嵌入在车辆中，而且它也是一个专业组件，所以它不太可能是人工智能即服务。</p> <p>c) 正确。独家算法不能是 AIaaS，尽管定价部分是通用的，可以作为人工智能即服务提供。</p> <p>d) 不正确。所描述的情况表明，该系统可以通过 web 提供，但只能在内部提供给他们自己的汽车租赁办公室，因此它看起来不像是人工智能即服务。</p>	AI-1.7.1	K2	1
5	b	<p>a) 不正确。SOTIF 是针对道路车辆的，不是针对潜艇的。</p> <p>b) 正确。银行贷款决策系统将使用个人数据，这是通用数据保护条例所涵盖的内容。</p> <p>c) 不正确。一辆完全自动驾驶的汽车可能包含非确定性系统，这是 ISO26262 所不允许的。</p> <p>d) 不正确。无人机碰撞系统不太可能包含任何个人数据，而这正是通用数据保护条例的重点。</p>	AI-1.9.1	K2	1
6	d	<p>a) 不正确。在无监督学习中，系统期望从无标记的数据中学习，而适应性作为一种特征与系统被改变有关，而不是改变系统本身。</p> <p>b) 不正确。有监督的学习依赖于有标签的数据，但灵活性与阅读没有标签的数据无关。</p> <p>c) 不正确。强化学习系统被期望优化奖励功能，而适应性作为一种特征，是与一个可以改变的系統相关联的，而不是一个自我改变的系统。</p> <p>d) 正确的。灵活性是一种与系统在原始需求之外的环境中使用的能力相关的特性。</p>	AI-2.1.1	K2	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
7	d	<p>a) 不正确。在相同的操作环境中工作的自学习系统仍然可以优化自己，从而改变自己的行为。</p> <p>b) 不正确。以人工智能为基础的系统，为了适应环境的变化而改变自己，仍然会产生负面的副作用。</p> <p>c) 不正确。演化作为一种特征并不是关于系统开发，而是关于基于人工智能的系统在其操作环境中部署后如何变化。</p> <p>d) 正确。如果自学习系统与人互动，那么它对自身做出的任何改变都有可能对人有潜在危害。</p>	AI-2.3.1	K2	1
8	c	<p>a) 不正确。这个游戏是针对儿童这个弱势群体的，可能会让他们对民主有不好的看法。</p> <p>b) 不正确。该申请可能会从社交媒体上获取私人人和/或有偏见的信息，这会对求职者的工作前景产生不利影响。</p> <p>c) 正确。尊重人的自主性、防止伤害、公平和可解释性等伦理原则不太可能受到水下海洋测绘系统的影响。</p> <p>d) 不正确。这样的应用程序有可能对弱势群体(如残疾人)不公平，还可能给员工带来不必要的压力。</p>	AI-2.5.1	K2	1
9	c	<p>a) 不正确。理解底层技术被认为是可理解性而不是可解释性。</p> <p>b) 不正确。理解贷款系统是如何做决定的，更容易解释。</p> <p>c) 正确。了解肿瘤学系统的基础技术可能是可理解的。</p> <p>d) 不正确。系统反应正确只是指功能上的正确性--而不是透明度、可理解性或可解释性。</p>	AI-2.7.1	K2	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
10	d	a) 不正确。在无监督的机器学习中，没有标签。 b) 不正确。强化学习系统使用一个适应度来驱动其学习。 c) 不正确。一个自我学习的强化系统使用奖励来教自己。 d) 正确。在无监督的机器学习中，没有标签，模式是由数据本身决定的。	AI-3.1.2	K2	1
11	b	考虑到给定的描述。 I. 证数据检查模型性能。这是在模型评估期间进行的。 II. 用于测试模型的测试数据的来源被确定。这是作为数据准备的一部分进行的。 III. 调整后的模型为其目标硬件做好准备。这是作为部署的一部分进行的。 IV. 测试数据用于确保满足商定的性能标准。这是作为测试模型的一部分而进行的。 V. 模型是由源代码创建的。这是作为建立和编译模型活动的一部分进行的。 VI. 关键数据特征被确定。这是作为数据准备的一部分进行的。 a) 不正确 b) 正确 c) 不正确 d) 不正确	AI-3.2.1	K2	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
12	a	a) 正确。不太可能在移动设备上训练 ML 算法。 b) 不正确。如果健康监测系统被植入并需要重新训练，我们可能会为其设定一个最长的时间，因为病人必须参加一个设施来进行训练，或者在远程进行训练时不受监测。 c) 不正确。预计模型使用的特征数量将影响模型的选择。 d) 不正确。在选择机器学习模型时，知道用于聚类的类的数量是一个有用的输入。	AI-3.4.1	K2	1
13	d	a) 不正确。测试数据集通常与验证数据集来自同一来源。 b) 不正确。测试数据集和验证数据集应该有相同的格式。 c) 不正确。测试数据集既不用于训练也不用于验证。 d) 正确。在训练中加入测试数据集会导致对模型的评估出现偏差。	AI-4.2.1	K2	1
14	c	a) 不正确。拥有不充分的数据并不能决定数据误标的可能性 b) 不正确。合成的数据不会导致不良的标签。 c) 正确。翻译错误可能导致一种语言中正确标注的数据在第二种语言中被错误标注。 d) 不正确。数据标签与机器学习算法的选择没有关系	AI-4.5.2	K1	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
15	a	<p>a) 正确。模型的质量取决于标签的质量。错误的标签会导致错误的真实数据。对于标签不正确的数据，功能性能测量可能表明模型质量很好，但它会产生错误的输出。</p> <p>b) 不正确。机器学习功能功能表现度量的值不依赖于用于测量它们的工具。</p> <p>c) 不正确。验证数据中可能有也可能没有偏差。</p> <p>d) 不正确。经常进行数据转换，它不一定影响模型的质量。错误的转换可能会导致数据质量问题 and 随后的模型质量问题，但不能得到一个与数据转换和模型质量差有关的一般性结论。</p>	AI-5.3.1	K2	1
16	a	<p>a) 正确。基准套件表明训练时间。</p> <p>b) 不正确。基准测试套件并不表明测试时间。</p> <p>c) 不正确。基准套件并不表明验证时间。</p> <p>d) 不正确。基准套件并不表明部署时间。</p>	AI-5.5.1	K2	1
17	b	<p>a) 不正确。组件测试适用于非模型组件，属于常规测试。</p> <p>b) 正确。进行输入数据测试是为了测试偏差，以及其他方面。</p> <p>c) 不正确。虽然独立偏差测试也可以作为系统测试的一部分，但理想情况下，它应该在训练模型之前完成。</p> <p>d) 不正确。模型测试用于检查模型是否单独满足任何指定的要求，例如，机器学习功能表现准则和非功能标准。</p>	AI-7.2.1	K2	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
18	c	a) 不正确，因为非功能需求是文件的一部分。 b) 不正确，因为接口的文档支持组件交互的白盒测试。 c) 正确，因为根据其来源和元数据对数据进行偏见测试是可能的。 d) 不正确，因为自适应的人工智能系统很少提供它们对自己所做的改变的文件。	AI-7.5.1	K2	1
19	c	a) 不正确，因为人工智能效应指的是随着时间的推移，人们对人工智能的看法发生了变化，而不是指某一解决方案的准确性。 b) 不正确。对抗性攻击是指攻击者巧妙地扰乱传递给训练好的模型的有效输入，使其提供不正确的预测。在这种情况下，数据是从购买历史中收集的，而不是来自用户的直接输入。因此，这是一个对抗性攻击的机会很低。 c) 正确。这是一个由于顾客行为的变化而使模型提供的准确性降低的例子。 d) 不正确。公平是指使用有积极偏见的数据进行训练，这在本案例中是不正确的，因为我们面对的是一个在过去表现良好的实时系统。	AI-7.6.1	K2	1
20	a	a) 正确。这与测试系统的自主性有关，因为可能需要在特定的时间段内进行测试，以检查系统需要干预的频率。 b) 不正确。一个系统可以是自主的，也可以是非常不准确的。 c) 不正确。一个系统可以是自主的，也可以是不适应的。 d) 不正确。训练数据的静态分析不太可能与自主性的测试有关。	AI-8.2.1	K2	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
21	c	a) 不正确。通常可以指定期望结果，但这可能需要包括一个容限或进行多次测试。 b) 不正确。非决定论系统不一定有可解释性问题。 c) 正确。这是对非决定性系统的有效描述，因此，测试可能需要运行几次才能得到统计上有效的测试结果。 d) 不正确。偏见和非决定性系统之间没有关联。	AI-8.2.1	K2	1
22	b	a) 不正确。测试深层神经网络可解释性的方法，需要执行系统，而深层神经网络是复杂的，无法理解。 b) 正确。测试透明度不需要执行被测系统，因为它只需要将文件与实现进行比较。 c) 不正确。自动化偏向于要求执行提供结果，并决定用户如何处理这些结果。 d) 不正确。如果不执行被测系统，就无法验证其准确性。	AI-8.6.1	K2	1
23	b	a) 不正确。数据不会导致测试结果参照物，因为它并不意味着难以确定期望结果。 b) 正确。如果一个预测系统的基础真理是未知的，那很可能会导致测试结果参照物。 c) 不正确。自主性与测试口令无关。 d) 不正确。确定期望结果，一般不需要知道系统是如何实施的。	AI-8.7.1	K2	1
24	c	a) 不正确。背靠背测试，而不是 A/B 测试，被称为差异性测试。 b) 不正确。A/B 测试可用于简单和复杂的机器学习模型。 c) 正确。A/B 测试需要有明确的验收标准，以便在被测试的两个模型之间进行选择。 d) 不正确。A/B 测试对编写测试案例没有帮助。	AI-9.4.1	K2	1

问题编号 (#)	正确答案	解释/原因	学习目标 L0	K	分数
25	c	<p>a) 不正确。在虚拟测试环境中运行罕见的场景是有用的，特别是当在现实世界中运行测试的成本过高时。</p> <p>b) 不正确。在虚拟测试环境中运行货币市场交易系统的测试是明智的，因为我们不能在真实市场上运行这些测试，因为有可能会损失金钱。</p> <p>c) 正确。在虚拟环境中更快地运行水果采摘机将无助于确定水果在实时生产中是否未受损害，因为在虚拟环境中检查水果的损害将比在真实（非虚拟）环境中检查真实水果的损害要困难得多。</p> <p>d) 不正确。出于安全考虑，在虚拟测试环境中为自动驾驶汽车运行危险的测试场景是一种良好的做法。</p>	AI-10.2.1	K2	1
26	a	<p>a) 正确。 分类模型形式的图像识别被用来识别 GUI 中的对象。</p> <p>b) 不正确。 概率软件工程可用于估计某些事情发生的概率。</p> <p>c) 不正确。 基于搜索的软件工程技术被用于减少大的问题空间。 这些可能对猜测图像对象的位置很有用，但不能把它们识别为图像。</p> <p>d) 不正确。 聚类是一种算法，而不是一种 AI 软件工程技术。</p>	AI-11.1.1	K2	1
27	c	<p>a) 不正确。视觉测试使用图像，不使用对象 ID。</p> <p>b) 不正确。视觉测试使用图像，但不进行逐个像素的比较。它使用人工智能来做比较。</p> <p>c) 正确。视觉测试可以找到类似于人类测试员的重叠的用户界面元素。</p> <p>d) 不正确。即使在布局改变时，视觉测试也能发挥作用。</p>	AI-11.6.1	K2	1