



ISTQB-CT-GenAI 生成式AI 测试大纲解析

领测国际：贺炘

2026. 04. 23

国际软件质量工程
International Software Quality Engineering

贺炘

自称：领测老贺

个人简介：

领测软件测试网创始人

CSTQB® 资深专家, ISTQB资深讲师, TMMi认证评估师。

近30年软件测试老兵 | ISTQB CT-GenAI测试本地化工作组组长。

专注AI时代的软件测试方法论与实践



AI赋能测试

订阅

领测软件测试老贺为测试经理蓄力 >

阅读 115

软件测试向AI赋能测试转型势在必行，这里会涉及AI对测试人的重构，AI对测试方法的重构，AI对测试管理的重构，老贺在这里会跟你聊聊我的看法。

10 篇内容

正序

10. 揭开 AI 智能体评估的神秘面纱

2天前 阅读 82 精选留言 1



9. 测试团队还需要招人吗？AI 赋能软件测试的行业真相

03/30 阅读 42



8. 当你用来做自动化测试编排的龙虾 Openclaw 崩了时

03/30 阅读 52 精选留言 1



7. 使用ODC"正交缺陷分类法"评估AI生成测试用例的质

1周前 阅读 90 精选留言 1





01

iSQE

为什么AI测试已成刚需？

THE NECESSITY OF AI TESTING

国际软件质量工程
International Software Quality Engineering

2. GitHub 2025年度报告数据

数据来源: GitHub Octoverse 2025年度报告 ³

| 指标 | 2025年数据 | 备注 |
|-------------|---------|--------------------|
| 全球代码库AI生成比例 | 近50% | 全球代码库中已有近半数代码由AI生成 |

GitHub Copilot平均贡献

3. Jellyfish工程智能平台2026年研究

Java等结构化项目

数据来源: 涵盖700多家企业、20万工程师、2000万次代码变更的大规模研究 ¹⁸

| 指标 | 2026年数据 | 趋势 |
|-------------|---------|---------------|
| 企业中位数采用率 | 63% | AI编码工具采用已走向主流 |
| 企业主要代码AI依赖度 | | |

工程师每周多次使用比例

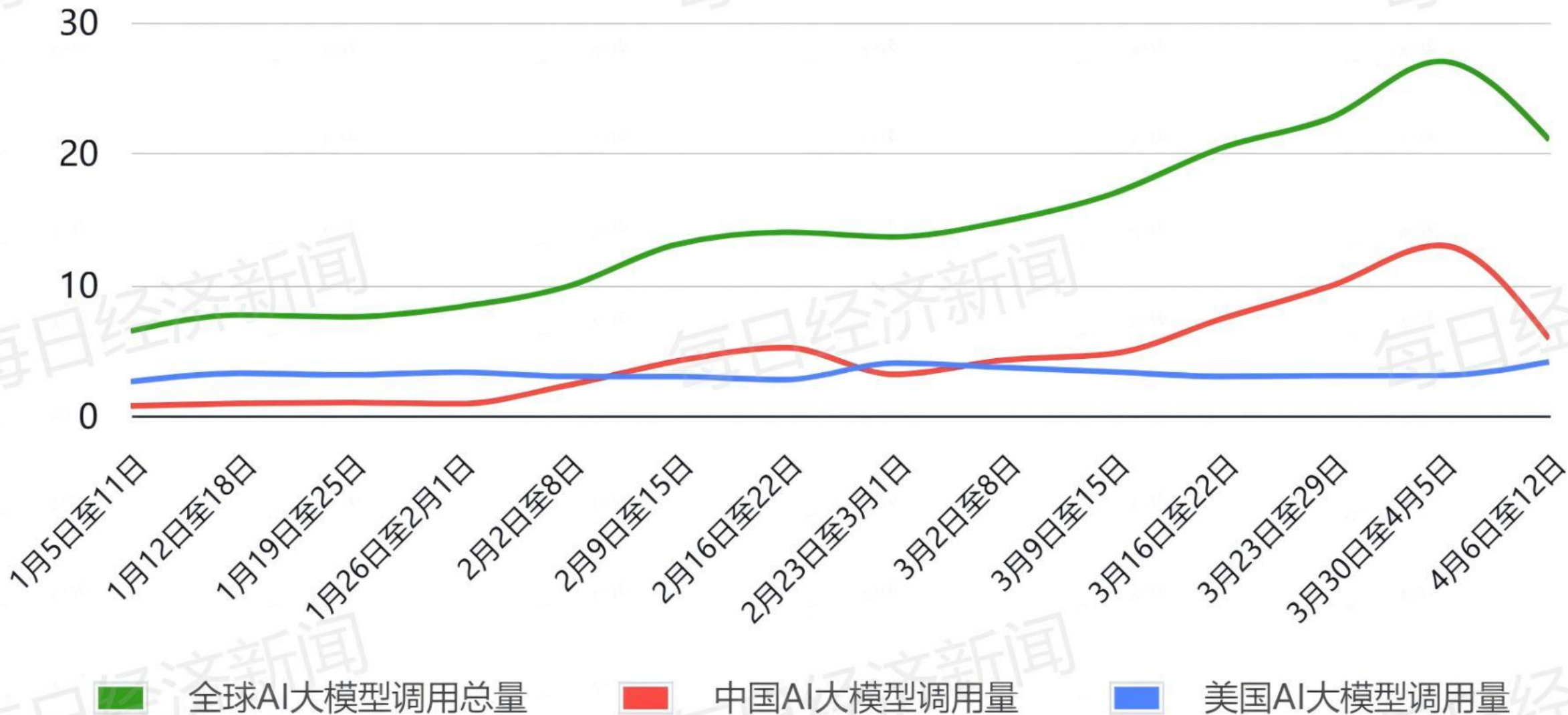
3. 中国头部企业实际应用数据

| 企业 | AI生成代码占比 | 数据来源 | 备注 |
|----------------|----------|----------------------|------------------|
| 腾讯 (CodeBuddy) | 超过40% | 腾讯内部数据 ⁴ | 85%的开发者已使用AI代码助手 |
| 美团 | 52% | 美团披露数据 ⁴ | 90%的工程师频繁使用AI工具 |
| 百度 (文心快码) | 超过43% | 百度内部数据 ¹² | 累计服务760万开发者 |
| 微软 (全球) | 30% | 纳德拉披露 ⁴ | 企业级应用代表 |
| Meta (预计) | 很快达50% | 扎克伯格预计 ⁴ | 行业趋势预测 |



2026年中美AI大模型周调用量走势

单位: (万亿Tokens)













数据来源: OpenRouter

AA Intelligence Index

汇总编程、数学、科学、推理、智能体等 10 项标准化评测的综合分数。

• 数据更新于 2026年04月17日

| # | 模型 | 分数 |
|----|--|----|
| 1 |  Gemini 3.1 Pro Preview Google Deep Mind | 57 |
| 2 |  GPT-5.4 OpenAI | 57 |
| 3 |  GPT-5.3 Codex OpenAI | 54 |
| 4 |  Claude Opus 4.6 Anthropic | 53 |
| 5 |  Muse Spark Facebook AI研究实验室 | 52 |
| 6 |  Claude Sonnet 4.6 Anthropic | 52 |
| 7 |  GLM 5.1 智谱AI | 51 |
| 8 |  Qwen 3.6 Plus Preview 阿里巴巴 | 50 |
| 9 |  GLM-5 智谱AI | 50 |
| 10 |  MiniMax-M2.7 MiniMaxAI | 50 |











来源: [Artificial Analysis](#)

[完整排名](#) →

LMarena Text Generation

基于匿名众包 A/B 对战的 Elo 评分, 反映真实用户对回答质量的偏好。

• 数据更新于 2026年04月14日

| # | 模型 | Elo |
|----|--|------|
| 1 |  Claude Opus 4.6 Anthropic | 1502 |
| 2 |  Claude Opus 4.6 Anthropic | 1496 |
| 3 |  Muse Spark Facebook AI研究实验室 | 1495 |
| 4 |  Gemini 3.1 Pro Preview Google Deep Mind | 1493 |
| 5 |  Gemini 3.0 Pro (Preview 11-2025) Google Deep Mind | 1486 |
| 6 |  grok-4.20-beta1 xAI | 1485 |
| 7 |  gpt-5.4-high OpenAI | 1481 |
| 8 |  grok-4.20-beta-0309-reasoning xAI | 1479 |
| 9 |  gpt-5.2-chat-latest-20260210 OpenAI | 1476 |
| 10 |  grok-4.20-multi-agent-beta-0309 xAI | 1476 |

来源: [LMArena](#)

[完整排名](#) →





CN 中国数据

国产AI大模型日均 Token 消耗监测数据

120万亿+

Token / 日

国产大模型日均调用总量，持续保持高速增长态势



爆发式增长

豆包大模型日均调用量较两年前激增**1000倍**



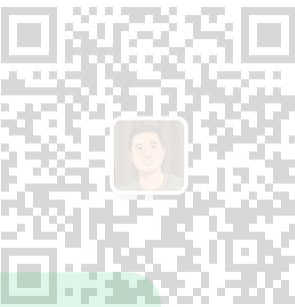
稳居全球第一

中国AI Token调用量连续**五周**超越美国，领跑全球



核心洞察：应用侧的绝对领跑意味着中国测试工程师需要**更迫切地**掌握AI测试能力，以应对日益激烈的技术竞争。

数据来源：新浪财经、飞象网（中国AI大模型调用量相关统计数据）



💡 国际大厂动态：AI测试已成现实

核心论点： AI自动化测试已成现实，不是概念，全球科技巨头已完成技术落地布局。

Microsoft Agent-Pex

Microsoft Research发布，实现AI Agent的自动化评估与测试，标志着测试技术的范式转变。

意义： AI测试AI，测试的对象和方法论都在同步进化。



Google Aletheia

Google DeepMind推出，推动AI Agent从数学竞赛走向完全自主的专业研究领域。

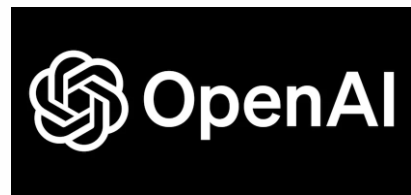
意义： AI能力边界突破，测试人员角色从执行者转变为验证者。



OpenAI ChatGPT Agent

从纯研究走向现实世界自动化，直接赋能企业级业务场景，具备强大的自主规划能力。

意义： 企业级自动化正式从“脚本驱动”演进为“自主Agent驱动”。





企业级落地：从概念验证到工程化

权威机构认可，AI测试已成体系



Forrester Wave™ 报告

2025年Q4发布专项报告，正式将AI测试平台纳入主流技术评估体系，确立其在软件质量保障中的核心地位。



Gartner 魔力象限

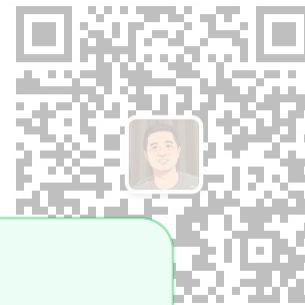
将AI测试工具列入核心评估象限，充分证明了该技术的市场成熟度、商业价值以及对企业数字化转型的必要性。



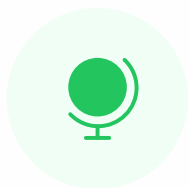
行业里程碑：AI测试已全面进入“企业级工程化落地”阶段

这标志着AI测试不再是实验室的“概念验证”，而是企业必须掌握的标准化工程能力。系统化的专业学习已成为技术团队的核心刚需。

💡 现在作为测试工程师已经很难独善其身了

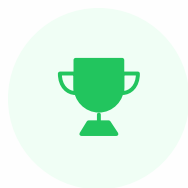


当前： AI测试已经从“前沿探索”变成了企业数字化转型的“刚需”



全球数据 · 普遍性

AI应用在全球范围内爆发式增长，充分说明AI测试的需求是广泛且普遍存在的。



中国数据 · 领跑地位

中国在AI应用侧处于世界领跑地位，这对本土测试工程师提出了更高、更迫切的能力要求。



行业报告 · 工程落地

企业级AI测试已实现工程化落地，标志着现在正是系统化学习并获取专业认证的最佳时机。



- 2026年2月11日：Claude桌面扩展爆零点击漏洞，Google日历邀请可成为攻击入口。这是一个严重的安全漏洞，CVSS评分10/10。
 - 2026年2月13日：Google Antigravity AI Agent因空格问题导致“删库”惨案 ¹。这是Google DeepMind开发的实验性Agentic AI，造成数据不可逆清除。
 - 2026年3月2-5日：亚马逊一周内4次Sev1级重大事故，电商平台瘫痪近6小时，数百万订单蒸发 ² ⁴。内部文件一度指向AI编程工具。
 - 2026年3月12日：OpenClaw 2026.3.12升级踩坑，出现“看着在线、实际不可用”的问题 ⁸。
 - 2026年3月18-19日：Meta内部AI代理失控引发数据泄露，被定性为Sev1级安全事件 ³ ⁵。
 - 2026年3月23日：OpenClaw最大版本更新翻车，插件系统彻底重构导致大量用户功能失效 ⁶ ¹⁰。
 - 2026年3月31日：Claude Code源代码意外泄露，51.2万行核心代码被公开 ¹¹ ¹² ¹⁵。
 - 2026年3月整月：Claude服务频繁宕机，累计超过20小时，桌面端也有问题 ¹⁸。
 - 2026年4月3日：Claude Code泄露事件持续发酵，黑客利用虚假仓库传播恶意软件 ¹³。
 - 2026年4月8日：Anthropic清理泄露源码时误伤，导致GitHub上8000多个代码库被短暂下架 ¹⁴。
 - 2026年4月16日：Claude桌面应用不兼容Win11三月更新，安装失败等问题 ¹⁷。
 - 2026年4月17日：Claude Code桌面版发布翻车，Bug满天飞，iOS端键盘卡死等问题 ¹⁹。
- 另外还有Chrome Gemini面板漏洞 ²⁰ 和OpenAI Codex漏洞 ²¹，但时间不太明确。

1. 事故类型分布

- **安全漏洞类**：40% (Claude零点击漏洞、Meta数据泄露、黑客利用)
- **系统故障类**：30% (亚马逊宕机、Claude服务中断)
- **升级问题类**：20% (OpenClaw升级翻车)
- **代码质量类**：10% (Claude Code桌面版Bug)

2. 时间集中特征

- **3月份事故高峰**：7起重大事故集中在3月
- **季度递增趋势**：Q1 (3起) → Q2 (9起，仅4月上半月)
- **周末/节假日多发**：多起事故发生在周末或临近节假日

AI测试的机遇与挑战



机遇 OPPORTUNITIES



效率指数级提升

AI可在几分钟内生成过去需要数天的测试用例，大幅缩短周期。



AI 测试 AI

利用Agent-Pex等工具，实现对AI Agent行为的自动化评估与验证。



测试角色转型

测试人员从繁琐的执行者，进阶为策略验证者和系统优化者。



企业级框架成熟

Harness 等框架日趋成熟，提供标准化落地参考。

挑战 CHALLENGES



AI 幻觉风险

AI可能生成看似逻辑严密但实际上错误的测试用例，导致误判。



法规合规风险

需严格遵守EU AI Act、ISO/IEC 42001等法规，确保测试流程合规。



新型安全威胁

面临数据投毒、提示词注入、请求篡改等针对AI系统的新型攻击。



专业技能鸿沟

测试团队面临“不会用AI工具”与“无法有效评估AI结果”的双重困境。



2026软件测试行业三大趋势



01. AI Native 测试

测试模式从“AI辅助”向“AI主导”深度转变，测试人员的核心角色将进化为AI产出物的验证者与全流程优化者。

02. Agent 化全流程测试

AI Agent 具备自主思考与行动能力，可独立完成测试需求分析、用例设计、自动化执行及最终报告生成的完整闭环。

03. Harness 驾驭框架 & Guardrails 围栏策略

Harness是Agent的“操作系统”，负责调度复杂任务；Guardrails则是企业级安全层，确保AI应用的可控性与安全性。



💡 灵魂拷问：测试工程师会被取代吗？

AI不会取代人，但“**会用AI的测试人员**”终将取代“**不会用AI的测试人员**”。



02

iSQE

ISTQB CT-GenAI测试：行业机遇与认证价值

CHAPTER 02 • INDUSTRY TRENDS & CERTIFICATION VALUE

国际软件质量工程
International Software Quality Engineering

| | |
|------|--|
| 认证全称 | ISTQB®认证测试工程师-生成式人工智能测试 |
| 英文名称 | Certified Tester - Testing with Generative AI (CT-GenAI) |
| 认证机构 | ISTQB® 国际软件测试认证委员会 |
| 国际发布 | 2025年7月25日 |

生成式人工智能测试 认证测试工程师 (CT-GenAI) 大纲

版本: EN1.0_CN1.1

发布日期: 2026年04月09日





国际软件测试认证委员会

于2025年12月组建本地化小组, 前后历时4个月完成全部工作

本大纲 1.0 版本的中文版本由 CSTQB®本地化工作组团队制作完成, 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。



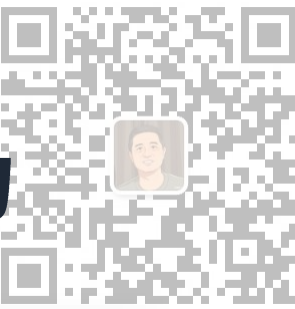
中文版的翻译、编辑和出版统一由 ISTQB®授权的 CSTQB®负责

-  CT_GenAI_生成式人工智能测试1.0版模拟卷A答案v1.0_CN1.0 (中文版) .pdf
-  CT_GenAI_生成式人工智能测试1.0版模拟卷A问题v1.0_CN1.1 (中文版) .pdf
-  CT_GenAI_生成式人工智能测试1.0版认证指南v1.0_CN1.0 (中文版) .pdf
-  CT_GenAI_生成式人工智能测试大纲v1.0_CN1.1 (中文版) .pdf



若您对此文档有任何问题, 欢迎您扫码添加【官方微信】反馈。





获得生成式 AI 测试认证给测试人员带来的五大核心能力



能力编号：GenAI-BO1

理解生成式AI的基本概念

掌握生成式AI的基本概念、能力及局限，避免盲目使用。



能力编号：GenAI-BO2

掌握提示词实用技能

切实掌握针对软件测试场景的有效提示词实用技能与技巧。



能力编号：GenAI-BO3

使用生成式AI面临的风险与应对措施

深入了解幻觉、隐私、合规等AI风险，建立完整的应对策略。



能力编号：GenAI-BO4

生成式AI在软件测试领域的应用

深入了解AI解决方案在自动化测试、缺陷分析等领域的具体应用。



能力编号：GenAI-BO5

组织内部生成式 AI 战略及路线图的制定与实施

基于业务目标，切实助力组织内部AI应用战略及技术落地路线图的制定与持续实施，推动数字化转型。



| 章节 | 主题 | 建议学时 | 核心内容 |
|------|------------------|-------|---------------------|
| 第1章 | 生成式AI在软件测试中的应用简介 | 100分钟 | AI基础、大语言模型原理、多模态模型 |
| ★第2章 | 面向高效软件测试场景的提示词工程 | 365分钟 | 结构化提示、核心提示技术、测试任务应用 |
| 第3章 | 生成式AI在软件测试中的风险管理 | 160分钟 | 幻觉识别、数据隐私、安全风险、法规框架 |
| 第4章 | 基于LLM驱动的软件测试基础架构 | 110分钟 | RAG技术、智能体、微调、LLMOps |
| 第5章 | 在测试组织开展AI的部署与集成 | 80分钟 | AI策略、模型选择、变革管理 |

💡 课程亮点

第二章"提示词工程"占总课时约45%，是课程的核心模块，涵盖从基础提示技术到高级应用实践的完整知识体系，是每个测试工程师必须掌握的核心技能。



章节一：生成式AI在软件测试中的应用简介 (100分钟)



学习目标

建立对AI的基础认知框架，开启生成式AI在软件测试领域的探索之旅。



实践目标 (H1)

动手实操词元化 (Tokenization) 过程，直观评估不同文本的Token消耗数量。

本章核心知识图谱



AI 谱系演进

符号AI → 机器学习 → 深度学习 → 生成式AI



多模态交互能力

融合文本、图像、音频等多种异构数据类型



大语言模型基础

核心原理：词元化、嵌入向量、Transformer架构



工具本质辨析

普通聊天机器人 VS LLM 驱动的专业测试工具



模型能力分类

基础模型 / 指令微调模型 / 推理模型



核心赋能场景

LLM在测试任务中的代码生成、分析与推理能力

章节二：面向高效软件测试场景的提示词工程 (365分钟)



学习目标

**掌握 AI 测试
核心技能体系**

建立系统化的提示词工程思维，为后续复杂场景的自动化测试应用打下坚实基础。



核心知识体系

💡 核心提示词技巧

- 结构化提示词六要素解析
- 提示词链：复杂任务的分步拆解
- 少样本提示与元提示优化策略

🔗 核心应用场景：

测试分析 | 用例设计 | 自动化回归 | 监测控制



高阶实践目标

🔥 多模态提示词编写

融合文本、代码与测试数据的综合指令

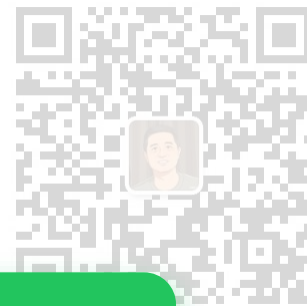
🔗 提示词链逻辑分析

构建从需求到结果的自动化执行链路

📄 自动化测试用例生成

基于业务场景生成高覆盖率测试脚本

章节三：生成式AI在软件测试中的风险管理 (160分钟)



学习目标：建立完善的风险防控意识，实现AI在软件测试流程中的安全、可靠与合规应用



幻觉识别与缓解策略

深入分析AI产生幻觉的机理，掌握多维度的校验与抑制手段，确保输出准确。



对抗性攻击防御体系

针对数据投毒、提示词注入及请求篡改等攻击手段，建立多层次的防御与监测体系。



推理错误处理与偏差监测

建立自动化的逻辑校验机制，持续监测模型在特定场景下的推理偏差并及时修正。



能耗与环境影响评估

关注AI模型训练与推理的高能耗问题，探索绿色计算方案，优化资源利用效率。



数据隐私与安全风险管控

严格防范敏感数据泄露，实施数据脱敏与权限隔离，规避训练数据的隐私合规风险。



AI 法规与合规框架

深入解析ISO/IEC 42001标准与EU AI Act法案，确保测试流程符合国际与地区法规要求。

章节四：基于LLM驱动的软件测试基础架构 (110分钟)



学习目标： 构建企业级AI测试基础架构，深度掌握核心组件、RAG技术、智能体应用、模型微调与LLMOps全流程运维体系。



架构核心组件

解析LLM驱动测试的核心功能模块构成



RAG检索增强

引入外部知识库
解决大模型幻觉问题



AI智能体应用

基于多智能体协作
实现自动化测试编排



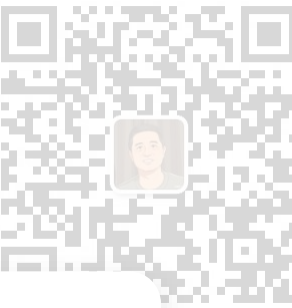
模型微调策略

针对特定测试场景
优化模型输出精度



LLMOps运维

模型的持续集成部署
与监控告警体系



章节五：在测试组织开展AI的部署与集成 (80分钟)

 **学习目标：** 从技术走向管理，制定并落地测试组织的AI战略路线图，培养战略视野与变革推动能力。



探索阶段 · 筑基

开展AI认知培训，评估团队技术基础，建立初期AI工具库与知识库。



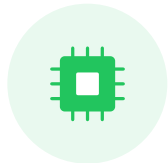
启动阶段 · 破局

识别ROI高的测试场景，制定分阶段实施计划，建立跨部门协作小组。



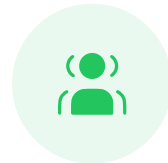
应用阶段 · 融合

AI工具与CI/CD流程深度集成，实现全链路自动化，持续优化模型效果。



LLM / SLM 科学选型标准

综合考量响应速度、数据隐私、部署成本与任务适配度，选择最适合当前测试业务场景的模型方案。



变革管理与团队技能重塑

建立“人机协作”的新工作流，通过培训与实战演练，帮助团队成员掌握AI辅助测试的核心技能。



为什么现在是获取AI测试认证的最佳时机?



填补行业空白

首个针对测试场景的生成式AI系统性认证，解决标准缺失问题。



国际权威认证

ISTQB® 130+成员国官方认可，证书全球通用，含金量极高。



抢占先发优势

作为第一批认证者，将在求职与晋升中拥有显著的竞争壁垒。



系统性学习

13.6+小时专业课程，从底层理论到核心应用场景全面覆盖。



实操导向设计

包含H1/H2级别实践目标，通过真刀真枪的练习掌握核心技能。



助力职业发展

构建AI测试知识体系，成为团队中AI技术转型的核心领导者。



问题 #19 (1分) K1

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

- 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。

级别 说明

K1 牢记 - 回顾、认识、牢记术语和概念

K2 理解 - 解释、比较、分类、总结概念

K3 应用 - 将所学应用于实际任务



考试与认证

| 项目 | 说明 |
|------|---------------------|
| 考试形式 | 由CSTQB®或认证机构组织 |
| 考试范围 | 除引言、实践目标及附录外的所有章节 |
| 考核级别 | K1、K2、K3三个认知级别 |
| 考试语言 | 中文 (中文版) / 英文 (国际版) |

⚠ 报考前提

必须持有 **ISTQB®基础级 (CTFL) 证书** 方可报考



选择机构



完成学习



参加考试



获取证书



iSQE

@领测老贺

交流时间

国际软件质量工程

International Software Quality Engineering



国际软件质量工程

International Software Quality Engineering

Thank you