

AI 驱动 软件研发 全面进入数字化时代

中国·深圳 11.24-25

AI+
software
Development
Digital
summit



基于AIGC构建具备中高级水 平的测试领域AI助手

于蕾 蚂蚁集团

科技生态圈峰会 + 深度研习



—1000+ 技术团队的选择



K+全球软件研发行业创新峰会

会议时间: 2024.05.24-25



K+全球软件研发行业创新峰会

会议时间: 2024.09.20-21



AI+ 软件研发数字峰会

会议时间: 2023.11.24-25



AI+ 软件研发数字峰会

会议时间: 2024.07.19-20



AI+ 软件研发数字峰会

会议时间: 2024.11.15-16

▶ 演讲嘉宾



于蕾

蚂蚁集团

蚂蚁集团平台工程与技术风险部质量技术测试开发专家，目前主要负责蚂蚁集团CIO部门的业务质量保障工作，包括蚂蚁集团业务风险、行政、采购、内控等多个业务领域相关的质量保障及稳定性，并负责测试分析大模型训练、领域模型评测等方向的相关工作，毕业后曾就职于百度，从事信贷业务后台、搜索等业务的质量保障

目录

CONTENTS

1. 蚂蚁测试流程与测试分析背景介绍
2. 测试分析模型训练介绍
3. 实践成果介绍
4. 展望

PART 01

蚂蚁测试流程与测试分析背景介绍

蚂蚁测试分析背景介绍

研发流程

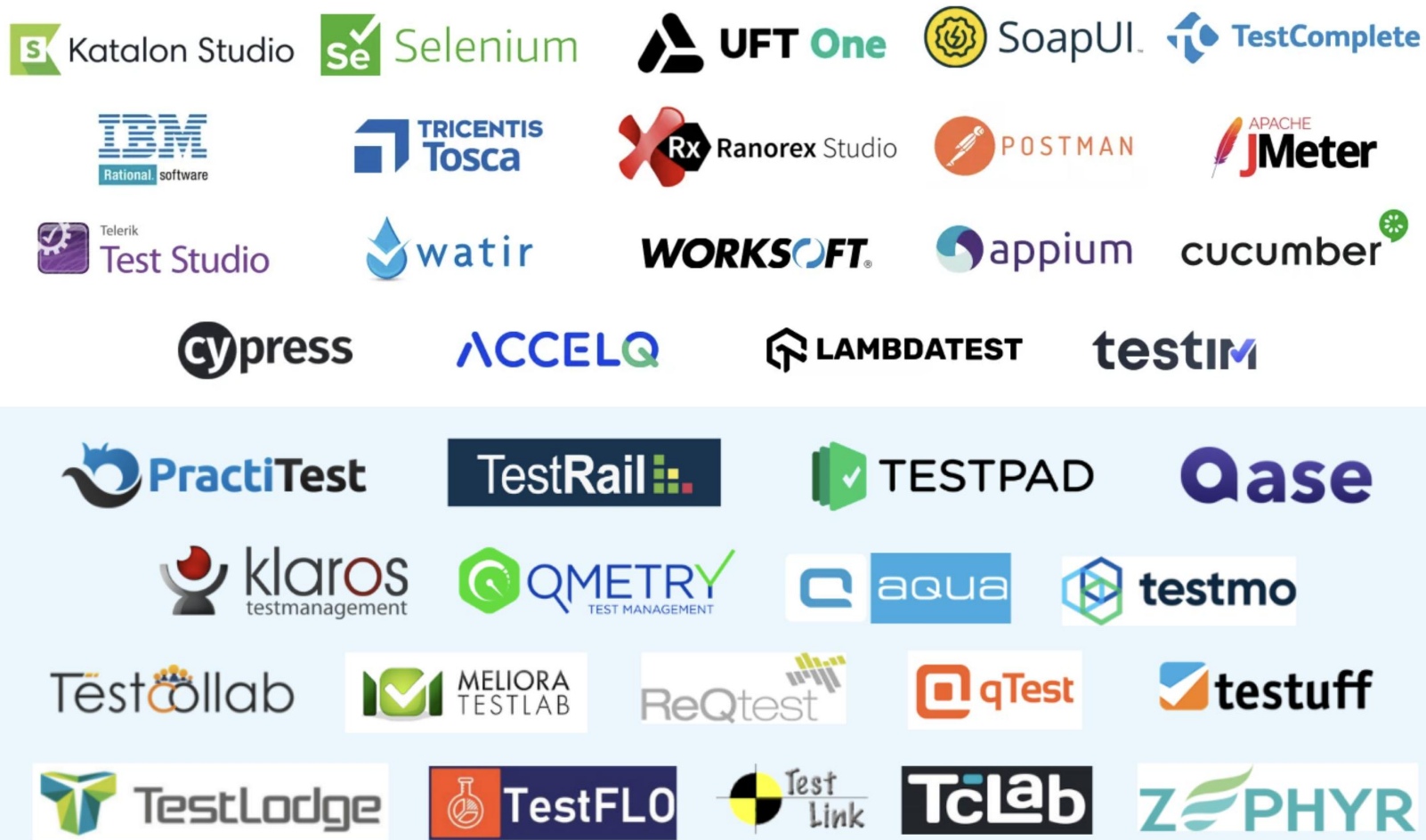


研发流程中测试分析环节是测试工程师质量保障工作最核心的价值产出



- 测试分析环节在整个研发流程提效难，针对大型项目通常需要**数天**才能完成；
- 测试人员能力与经验依赖性强，通常需要**1年以上**培养一个业务领域质量工程师；**2-3年**培养一个业务领域质量专家；

▶ 测试领域工具现状

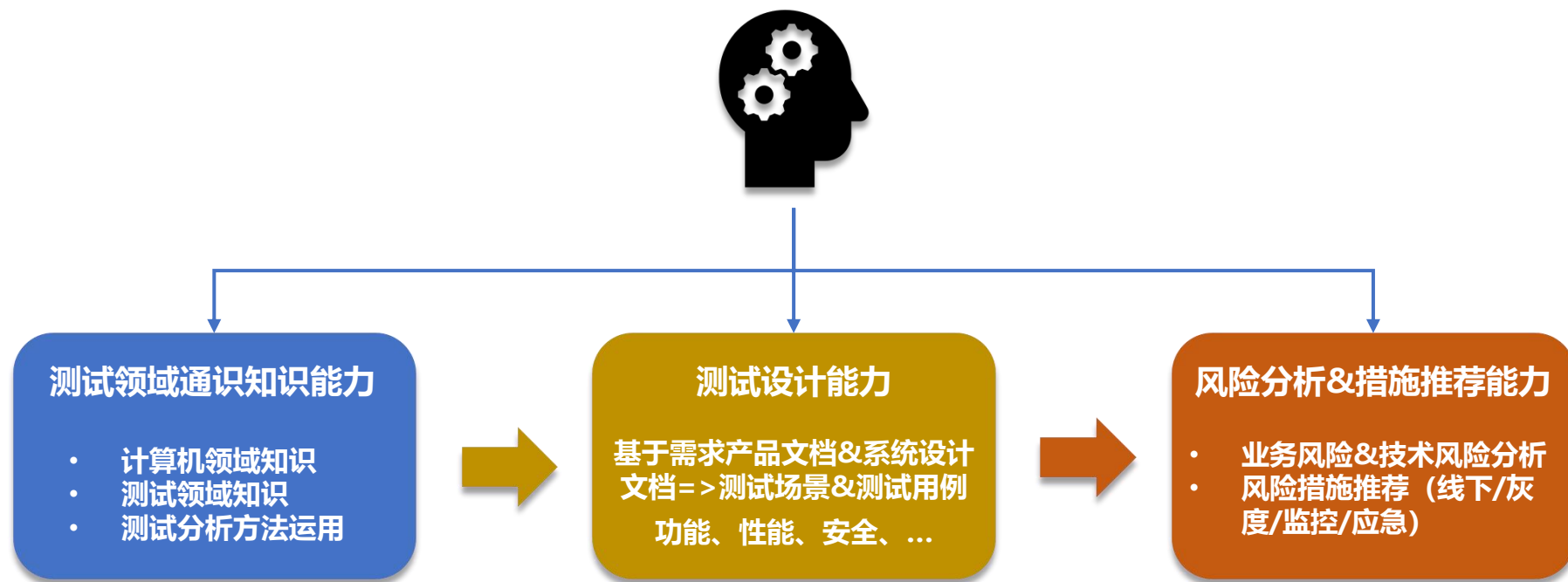


- 测试用例管理 ✓
- 测试框架 ✓
- App/UI/接口 ✓
- DevOps ✓
- 测试分析生成 ✗

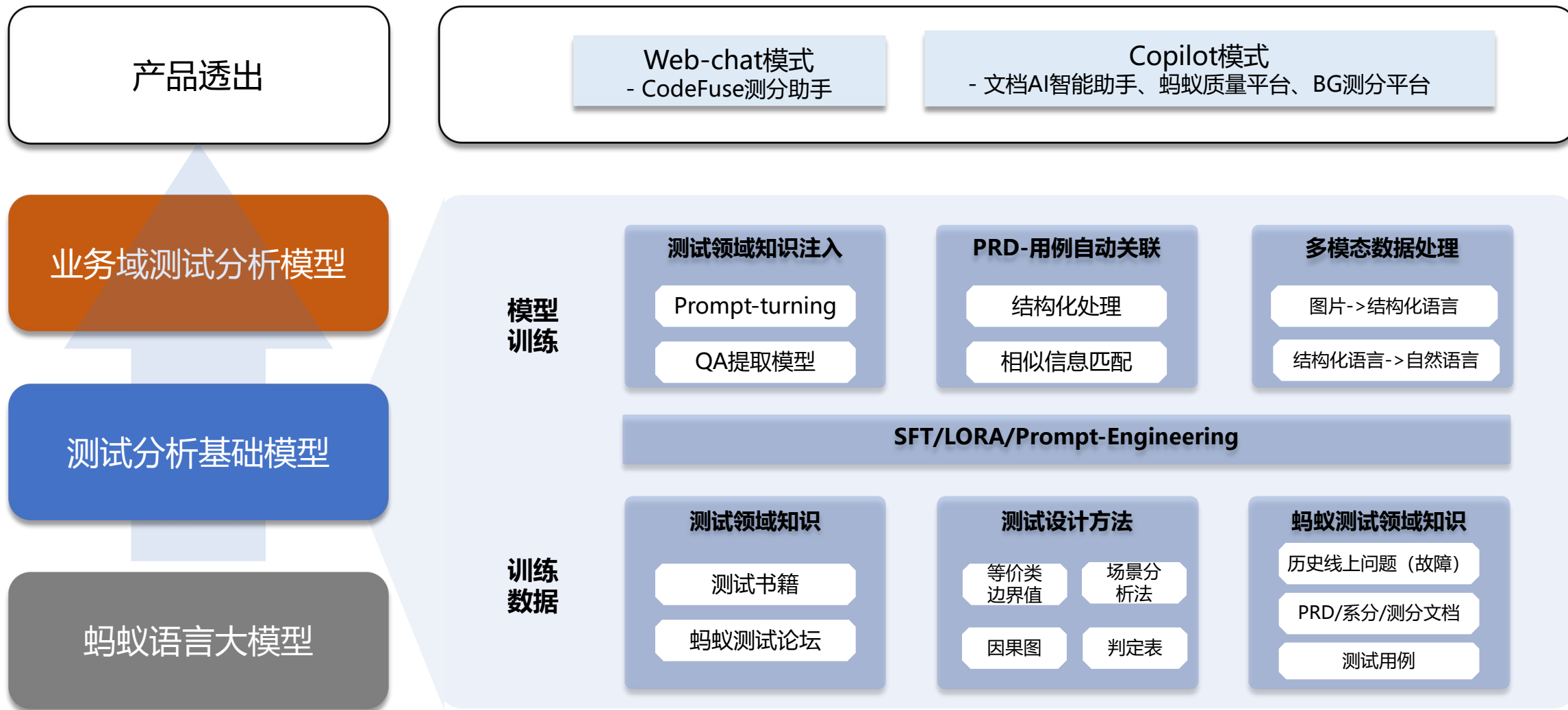
PART 02

测试分析模型训练介绍

▶▶ 测试分析模型目标 - 中高级水平的测试领域AI助手



测试分析模型 – 模型架构



测试分析模型 – 测试领域知识注入

原文

把测试对象看作一个打开的盒子。利用白盒测试法进行动态测试时，可通过测试来检测产品内部动作是否按照规格说明书的规定正常进行，按照程序内部的结构测试程序，检验程序中的每条通路是否都能按预定要求正确工作，而不顾它的功能，白盒测试的主要方法有逻辑驱动、基路测试等，主要用于软件验证。白盒测试法的覆盖标准有逻辑覆盖、循环覆盖和基本路径测试。其中逻辑覆盖包括语句覆盖、判（断）定覆盖、条件覆盖、判（断）定/条件覆盖、条件组合覆盖和路径覆盖。这六种覆盖标准发现错误的的能力呈由弱至强的变化。\\n“白盒”法全面了解程序内部逻辑结构、对所有逻辑路径进行测试。“白盒”法是穷举路径测试。在使用这一方案时，测试者必须检查程序的内部结构，从检查程序的逻辑着手，得出测试数据。

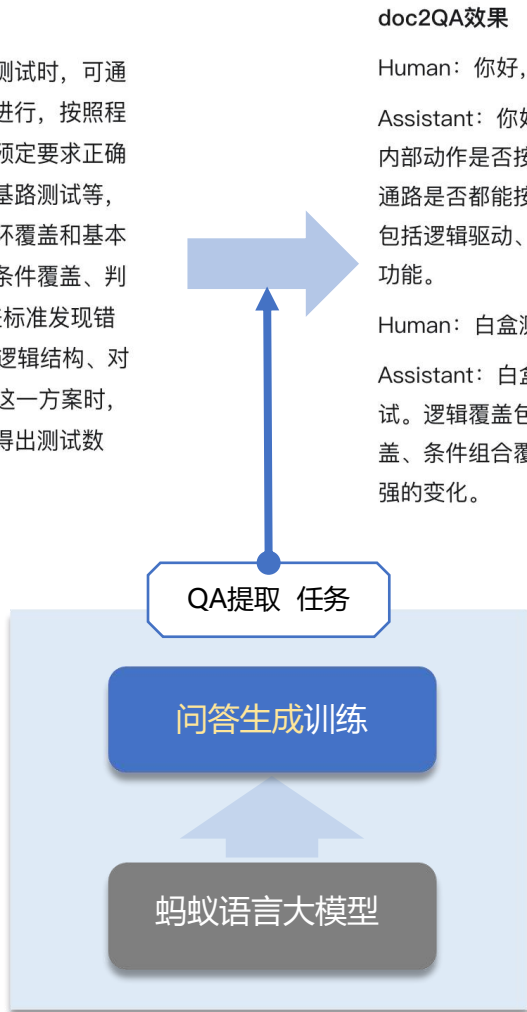
doc2QA效果

Human: 你好，请问白盒测试法是如何进行的？

Assistant: 你好，白盒测试法是一种动态测试方法，通过测试来检查产品内部动作是否按照规格说明书的规定正常进行。它主要测试程序中的每条通路是否都能按预定要求正确工作，而不顾产品的功能。白盒测试法主要包括逻辑驱动、基路测试等，可以涵盖程序内部结构、所有逻辑路径以及功能。

Human: 白盒测试法的覆盖标准有哪些？

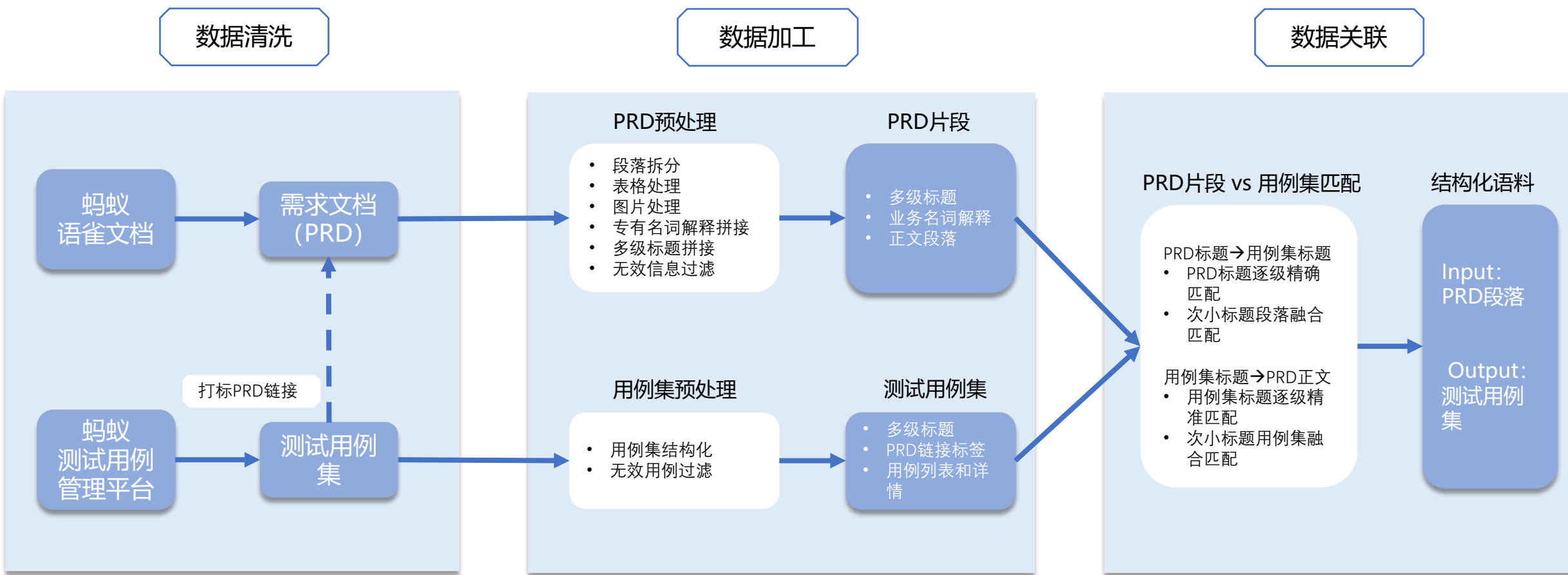
Assistant: 白盒测试法的覆盖标准包括逻辑覆盖、循环覆盖和基本路径测试。逻辑覆盖包括语句覆盖、判(断)定覆盖、条件覆盖、判(断)定/条件覆盖、条件组合覆盖和路径覆盖。这六种覆盖标准发现错误的的能力呈由弱至强的变化。



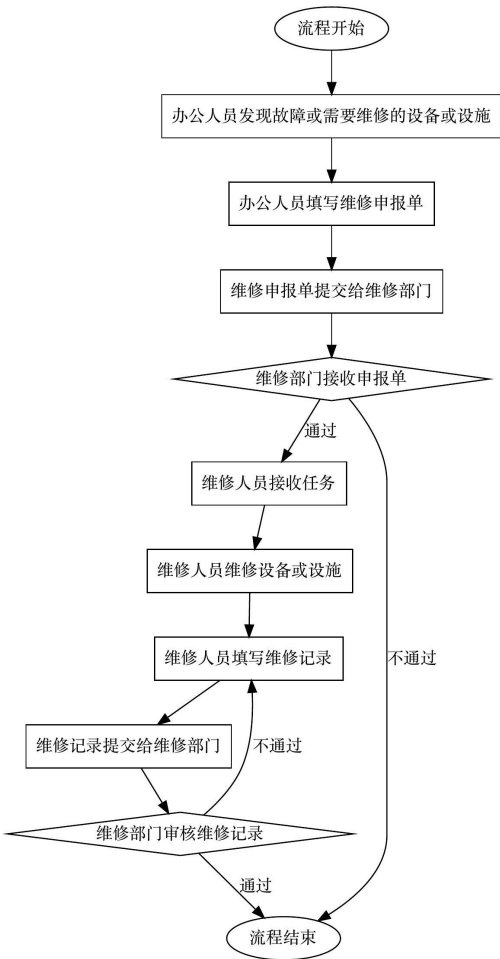
结构化语料构建

1. Input: Q
Output: A
2. Input: 原文 + Q
Output: A
3. Input: 相似Q
Output: 相似A
4. Input: QA + 相似Q
Output: 相似A

测试分析模型 – PRD-用例自动关联



测试分析模型 - 多模态数据处理



Step1

```
# Add nodes and edges
dot.node('0', '流程开始', shape='ellipse')
dot.node('1', '办公人员发现故障或需要维修的设备或设施', shape='box')
dot.node('2', '办公人员填写维修申报单', shape='box')
dot.node('3', '维修申报单提交给维修部门', shape='box')
dot.node('4', '维修部门接收申报单', shape='diamond')
dot.node('5', '维修人员接收任务', shape='box')
dot.node('6', '维修人员维修设备或设施', shape='box')
dot.node('7', '维修人员填写维修记录', shape='box')
dot.node('8', '维修记录提交给维修部门', shape='box')
dot.node('9', '维修部门审核维修记录', shape='diamond')
dot.node('10', '流程结束', shape='ellipse')
dot.edge('0', '1')
dot.edge('1', '2')
dot.edge('2', '3')
dot.edge('3', '4')
dot.edge('4', '5', '通过')
dot.edge('4', '10', '不通过')
dot.edge('5', '6')
dot.edge('6', '7')
dot.edge('7', '8')
dot.edge('8', '9')
dot.edge('9', '10', '通过')
dot.edge('9', '7', '不通过')
```

Step2

这张图片是一个流程图，描述了一个维修设备或设施的过程。这个过程包括以下步骤：

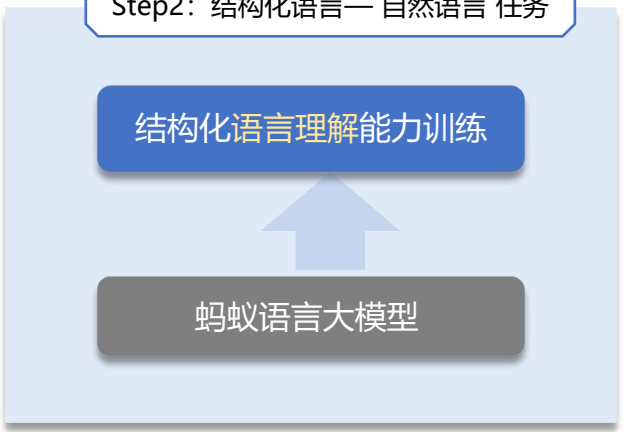
1. 流程开始。
2. 办公人员发现故障或需要维修的设备或设施。
3. 办公人员填写维修申报单。
4. 维修申报单提交给维修部门。
5. 维修部门接收申报单，维修人员接收任务并维修设备或设施。如果维修部门审核维修记录通过，进入步骤9；否则，返回步骤7。
6. 维修人员填写维修记录。
7. 维修记录提交给维修部门。
8. 维修部门审核维修记录。如果审核不通过，返回步骤7；否则，进入步骤10。
9. 流程结束，但设备或设施并未维修完成。
10. 流程结束，设备或设施已经维修完成。

每个步骤用一个节点表示，节点的形状和颜色可以根据需要进行自定义。箭头表示不同节点之间的关系，例如从节点2到节点3的箭头表示在“办公人员填写维修申报单”之后，“维修申报单提交给维修部门”。有些箭头上还带有文字说明，表示该关系的具体内容。整个流程的图示化，有助于管理者和执行者更好地理解 and 执行该维修流程。

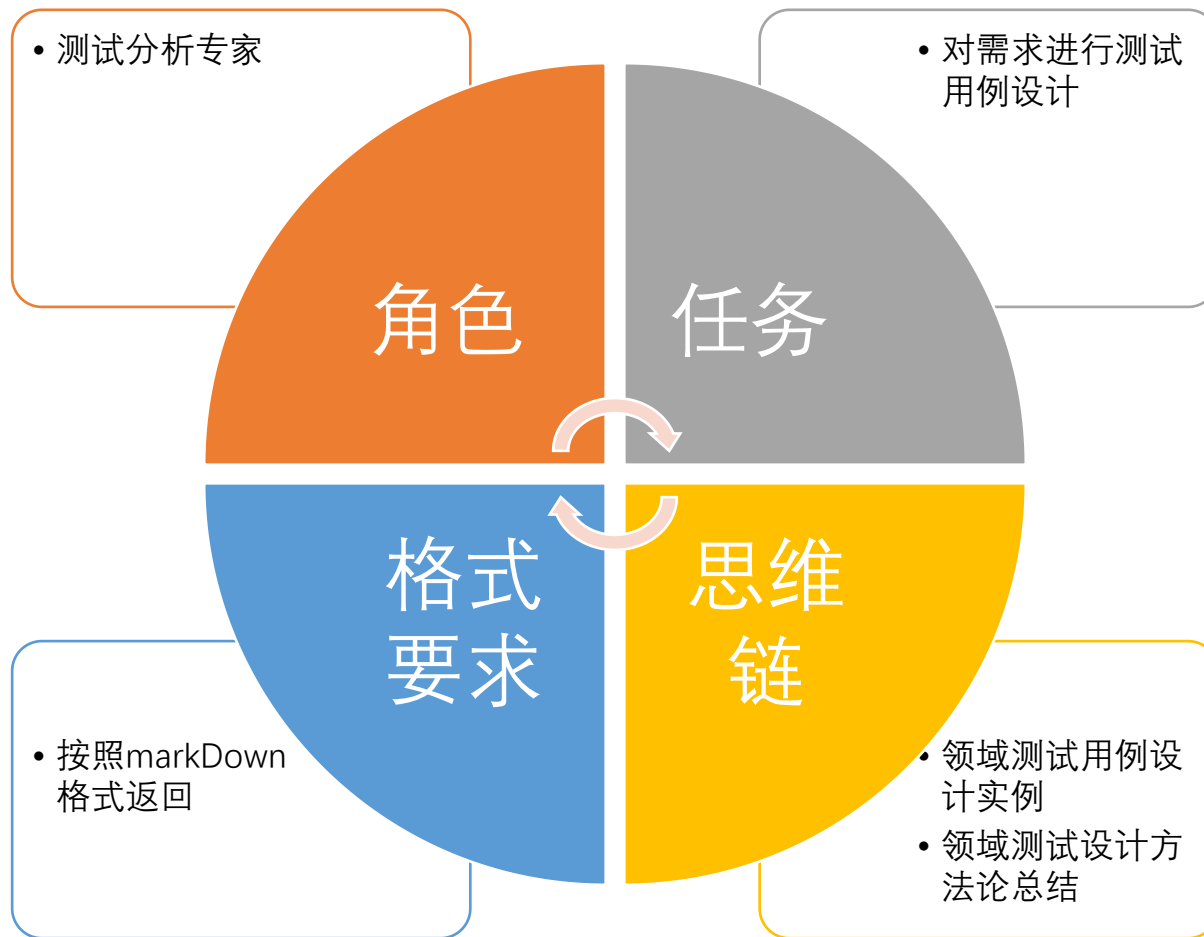
Step1: 图片—结构化语言 任务



Step2: 结构化语言—自然语言 任务



▶ 测试分析模型 – prompt设计



PART 03

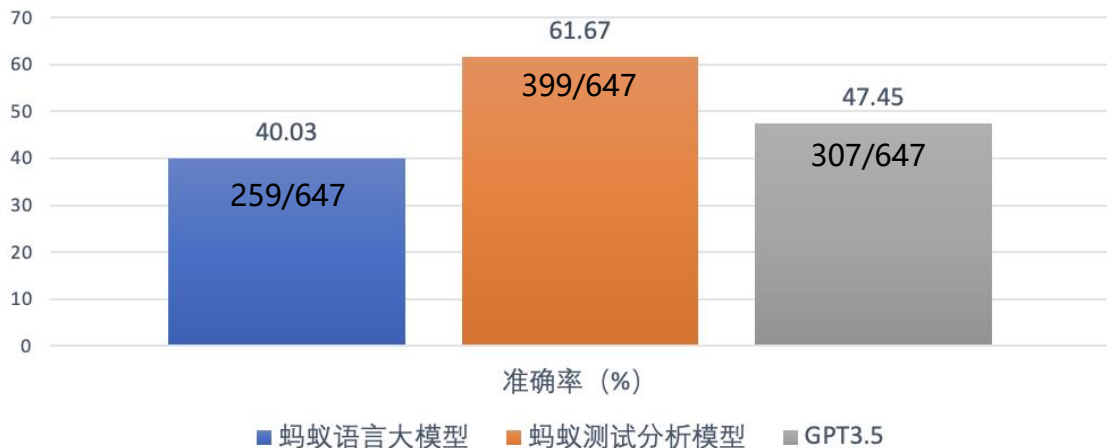
模型效果介绍



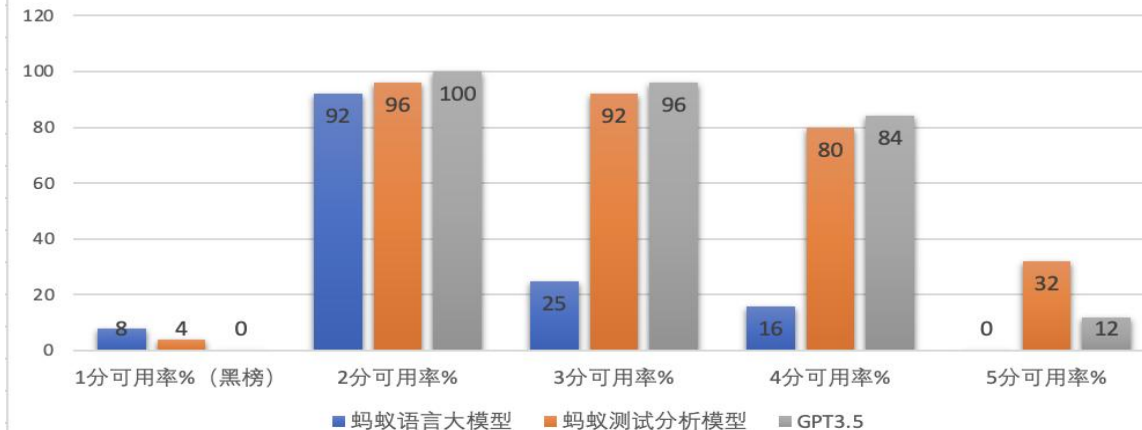
测模型效果 – 中高级软件考试

- 计算机技术与软件专业技术资格（水平）考试 – 中级/高级（软件评测师考试等）

中高级软考-客观题



中高级软考-主观题



e.g. 什么是白盒测试?

标准答案

白盒测试也称结构测试，透明盒测试。主要用于单元测试阶段，代码和逻辑的测试，重点复杂的测试，是一种测试用例设计方法，不同于黑盒测试，白盒测试是可以看到内部代码如何运作的，可通过测试来检测产品内部是否符合规定正常运行。

GPT3.5

白盒测试是一种软件测试方法，也称为结构测试或代码测试。它基于对软件内部结构和代码的了解，通过检查程序的内部逻辑、数据流、变量使用等来验证软件是否按照设计要求正确地执行。白盒测试通常由开发人员或专业测试人员执行，旨在发现代码中的错误、缺陷和漏洞，以提高软件质量和可靠性。

蚂蚁测试分析模型

白盒测试是一种软件测试方法，它涉及测试软件代码的内部实现，以验证其是否满足设计规格说明。白盒测试通常用于测试软件的内部逻辑和算法，以及检查代码覆盖率和分析代码性能。白盒测试的目标是验证软件是否满足规格说明，并检测和修复潜在的错误和缺陷。白盒测试通常包括静态分析、动态分析和逻辑覆盖率分析等技术。

模型效果 - 测试用例设计&风险措施分析

使用测试设计方法进行用例设计

Human: 假如你是一个经验丰富的测试工程师, QQ号码要求是6-10位的整数, 请使用等价类划分法进行测试用例设计

Answer: 根据等价类划分方法, 我们可以将QQ号码的输入范围划分为以下几个等价类:

- 6位及以下: 这个等价类包括最小长度和最大长度的情况。
- 6-10位: 这个等价类包括符合要求的6-10位数的范围。
- 10位以上: 这个等价类包括超过最大长度的10位数的范围。

基于以上等价类, 我们可以设计以下测试用例:

- 输入一个6位数的QQ号码, 例如"123456"。
- 输入一个10位数的QQ号码, 例如"12345678901"。
- 输入一个6-10位数的QQ号码, 例如"123456789012"。
- 输入一个超过最大长度的10位数的QQ号码, 例如"123456789012345678901"。

通过以上测试用例的设计, 我们可以覆盖到不同等价类的情况, 以验证QQ号码输入的合法性。

根据业务场景生成测试分析

Human: 假如你是一个经验丰富的测试专家, 请针对电商平台网站的用户注册功能进行测试分析设计

Answer: 测试分析设计是测试过程中的重要环节, 它帮助测试人员确定测试的范围、目标和方法。针对电商平台网站的用户注册功能, 以下是一个测试分析设计的示例:

确定测试目标:

验证用户注册功能是否正常工作。

确保用户能够成功注册账号。

检查注册过程中的输入验证和错误处理机制。

确保注册后的账号信息能够正确保存和显示。

确定测试范围:

注册页面的各个输入字段 (如用户名、密码、邮箱等)。

注册过程中的各个步骤 (如填写必填信息、验证输入等)。

注册成功后账号信息的保存和显示。

设计测试用例:

正常注册: 输入有效的用户名、密码、邮箱等信息, 验证是否成功注册账号。

注册必填信息: 不填写必填信息, 验证是否能够正确提示用户必填信息。

注册格式验证: 输入特殊字符、过长或过短的邮箱地址, 验证是否能够正确验证并给出错误提示。

注册重复用户名: 输入已存在的用户名, 验证是否能够正确提示用户已存在。

注册邮箱验证: 输入不存在的邮箱地址, 验证是否能够正确提示用户邮箱不存在。

注册密码强度: 输入弱密码 (如长度小于8位、包含常见密码等), 验证是否能够正确提示用户密码强度不足。

注册成功后的账号信息: 验证注册成功后账号信息是否正确保存, 并能够正确显示。

根据业务场景分析风险防控策略

Human: 蚂蚁的三板斧指的是"可监控"、"可灰度"、"可应急"三方面的能力, 请根据下面的需求, 分析这个变更的三板斧策略以保障业务线上可用: 电商平台网站的用户注册功能

Answer: 根据您提供的信息, 针对电商平台网站的用户注册功能, 可以考虑以下三板斧策略以保障业务线上可用:

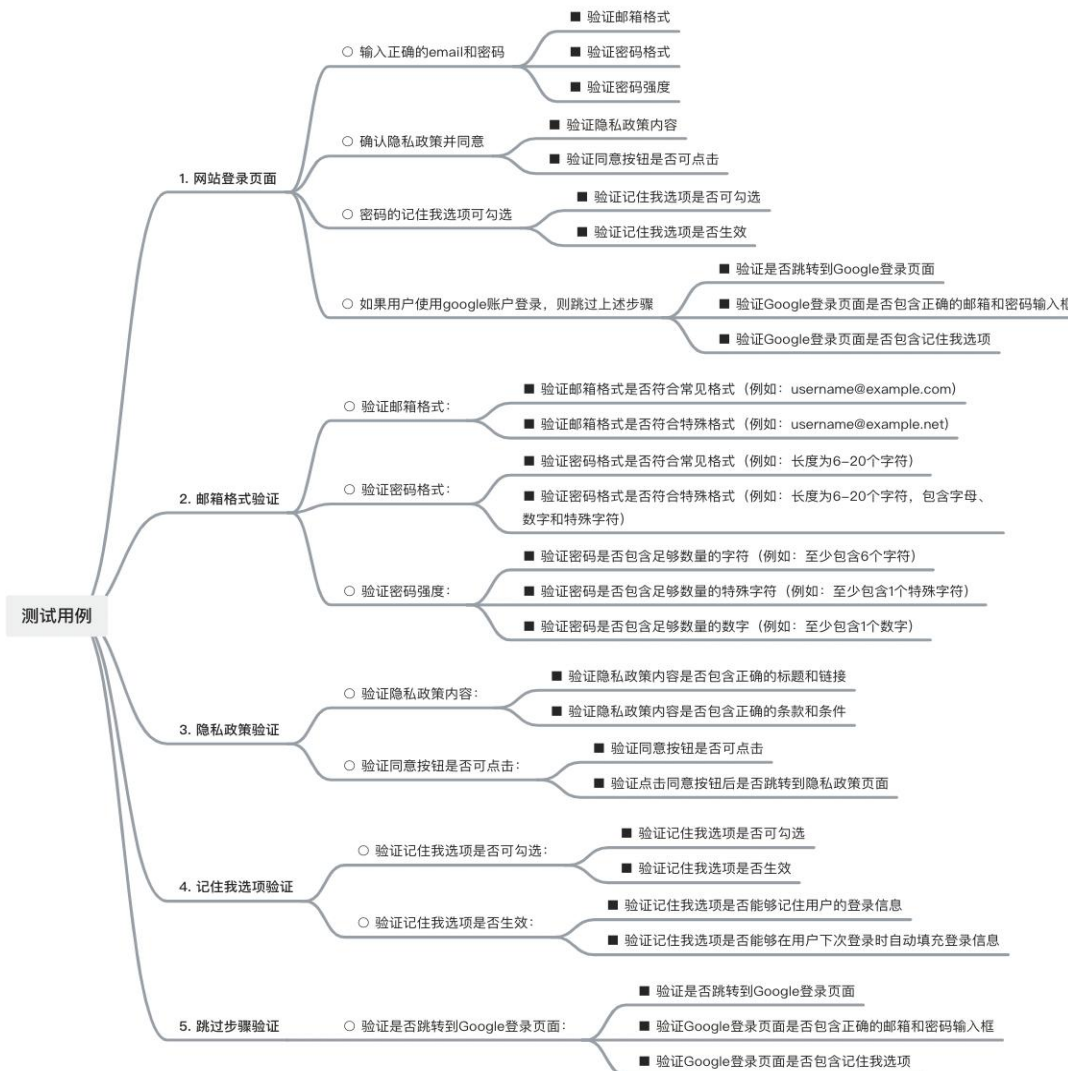
- 可监控:** 在用户注册功能上线后, 可以通过监控接口调用情况、注册成功/失败率、用户信息一致性等指标, 及时发现线上问题并快速响应。
- 可灰度:** 在用户注册功能上线后, 可以先进行小范围灰度验证, 例如先开放给部分用户进行注册, 再逐步扩大灰度范围, 确保线上业务稳定运行。
- 可应急:** 在用户注册功能出现线上问题时, 可以通过应急开关回切、代码回滚、用户入口屏蔽等策略, 快速恢复线上业务, 降低故障影响。同时, 在灰度验证过程中, 也可以收集用户反馈, 及时优化功能, 提高用户体验。

需要注意的是, 在实施三板斧策略时, 需要根据实际情况进行灵活调整, 例如在灰度验证过程中, 可以根据用户反馈和数据指标, 逐步扩大灰度范围, 提高灰度效果。同时, 也需要建立完善的应急响应机制, 确保在出现线上问题时能够快速响应并恢复业务。

模型效果 - 文档AI智能助手

PRD片段:

网站登录页面，需要输入正确的email和密码，同时需要确认隐私政策并同意。密码的『记住我』选项可勾选，如果用户使用google账户登录，则跳过上述步骤



PART 04

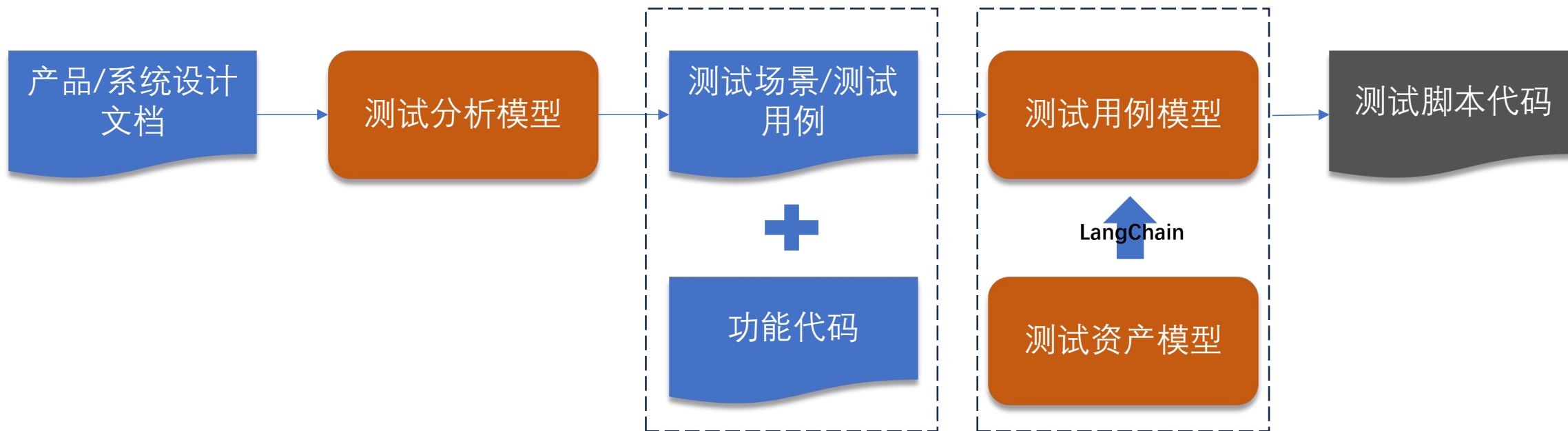
展望



▶ 展望 - 质量专家



▶ 展望 – 从测试分析到测试代码生成



THANKS

