



2025 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发

05/23-24 | 上海站





2025 AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

08/08-09 | 北京站

11/14-15 | 深圳站



查看会议详情

北京站论坛设置

大模型和 AI 应用评测

智能存储与检索技术

下一代知识工程

AI+ 金融业务创新

智能需求工程

智能体与研发效率工具

AI 产品运营与出海策略

大模型安全与对齐

大模型应用开发框架与实践

智能体经济 (Agentic Economy)

智能测试工具的开发与应用

具身智能与机器人

代码生成及其改进

AI+ 新能源汽车

AI 前沿技术探索与实践

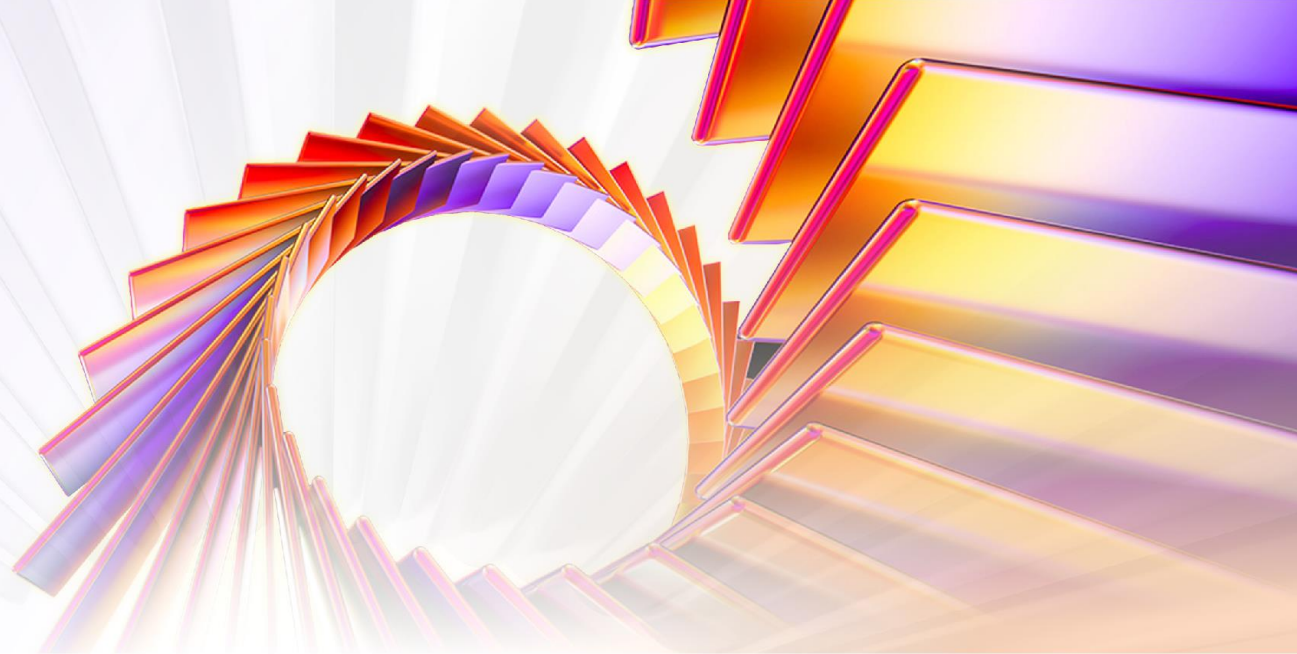


| 05/23-24 | 上海站

2025 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



AI驱动的需求洞察与自动化 闭环实践

彭华聪、傅汤毅 | B站



彭华聪

B站资深测试开发工程师

B站供给生态质量团队资深测试开发工程师。目前主要负责B站内容审核业务质量体系建设和探索。也曾任职阿里巴巴测试开发专家，负责过搜索、实人认证等产品从事质量工作十余年，经验丰富。



傅汤毅

B站资深测试开发工程师

高级测试开发工程师，推进并落地了业务线自动化建设，流水线覆盖，有丰富的服务端经验及平台建设经验。

目录

CONTENTS

- I. 背景
- II. 问题/痛点
- III. 解决思路/整体方案
- IV. 具体实现/技术实践
- V. 总结与展望

PART 01

背景



工作量激增

软件迭代周期缩短1-2周，测试工作量极速增长，人工测试难以满足效率要求。



质量要求提升

分布式架构的复杂性不断增长，面临的技术环境。对于稳定性、体验、资损等专项类保障期望逐渐增多



成本压力显著

人力成本上涨与质量高要求得矛盾要求需要技术手段优化投入产出比





01

技术成熟

技术成熟度显著提升，各类AI模型准确率不断提升

02

基础完善

AI服务从模型种类、基础服务提供等方面得到广泛应用。接入门槛大大降低

03

广泛应用

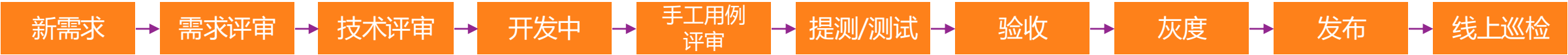
AI在自动驾驶、医疗等领域的应用取的可观的成效，为全面替代传统测试奠定技术基础。



PART 02

问题

▶ 传统迭代模式的弊端



人工成本大

用例
评审

测试用例人为编写需花费大量时间

需求
评审

需求排期采用班车制，需求集中处理，人工评审、排期调整工作量大

提测
测试

无法对需求新增或修改的接口自动化用例完成快速覆盖，人为新增需花费大量精力和时间

提测
测试

页面测试过程中需频繁变更账号或环境，手动操作各个环节，麻烦且费时

维护困难

线上
巡检

传统UI自动化用例测试方式为DOM扫描维护成本高

提测
测试

用例平台执行的接口自动化用例结果不稳定，维护投入大

测试稳定性有限

线上
巡检

传统UI自动化用例测试且易受前端迭代影响，不太稳定

容易遗漏

用例
评审

手工测试用例的一些临界点容易出现遗漏

提测
测试

自动化用例的边界情况无法测试完全

排查不方便

提测
测试

自动化用例的测试结果一旦出现报错，由于用例存在多个节点且有上下文依赖关系，难以快速排查出错误原因

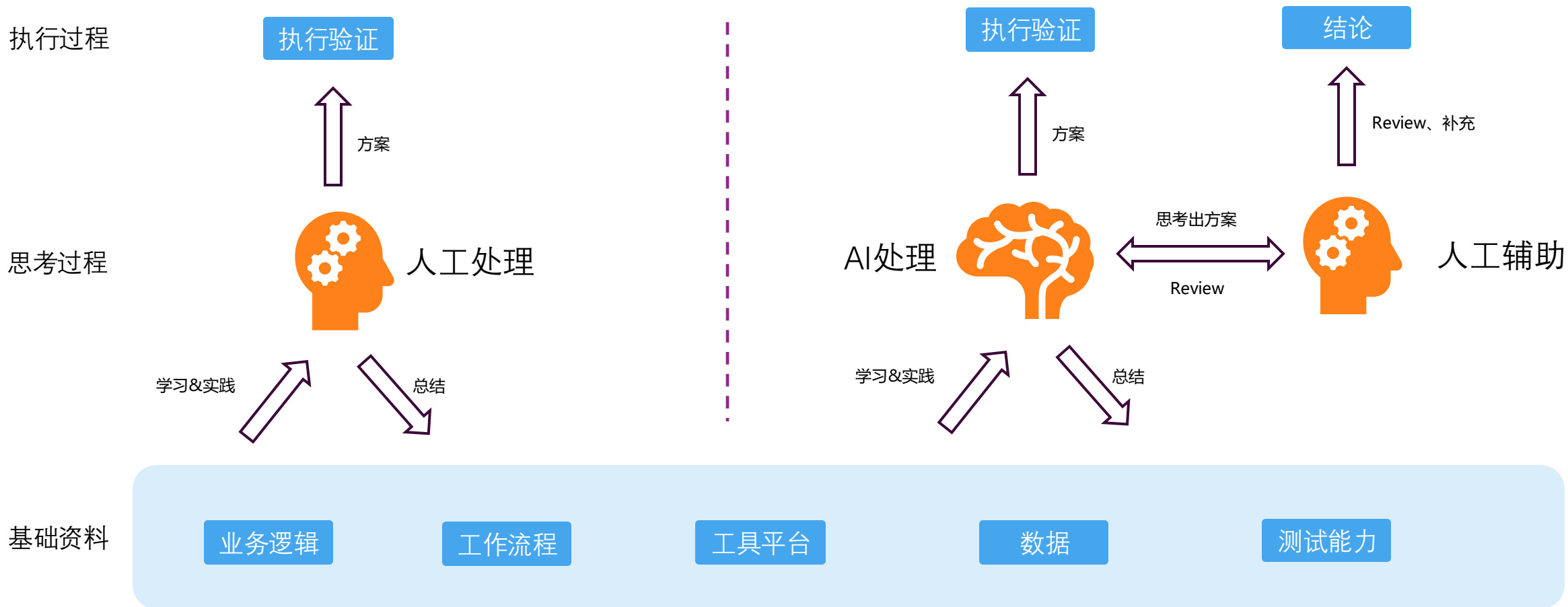


PART 03

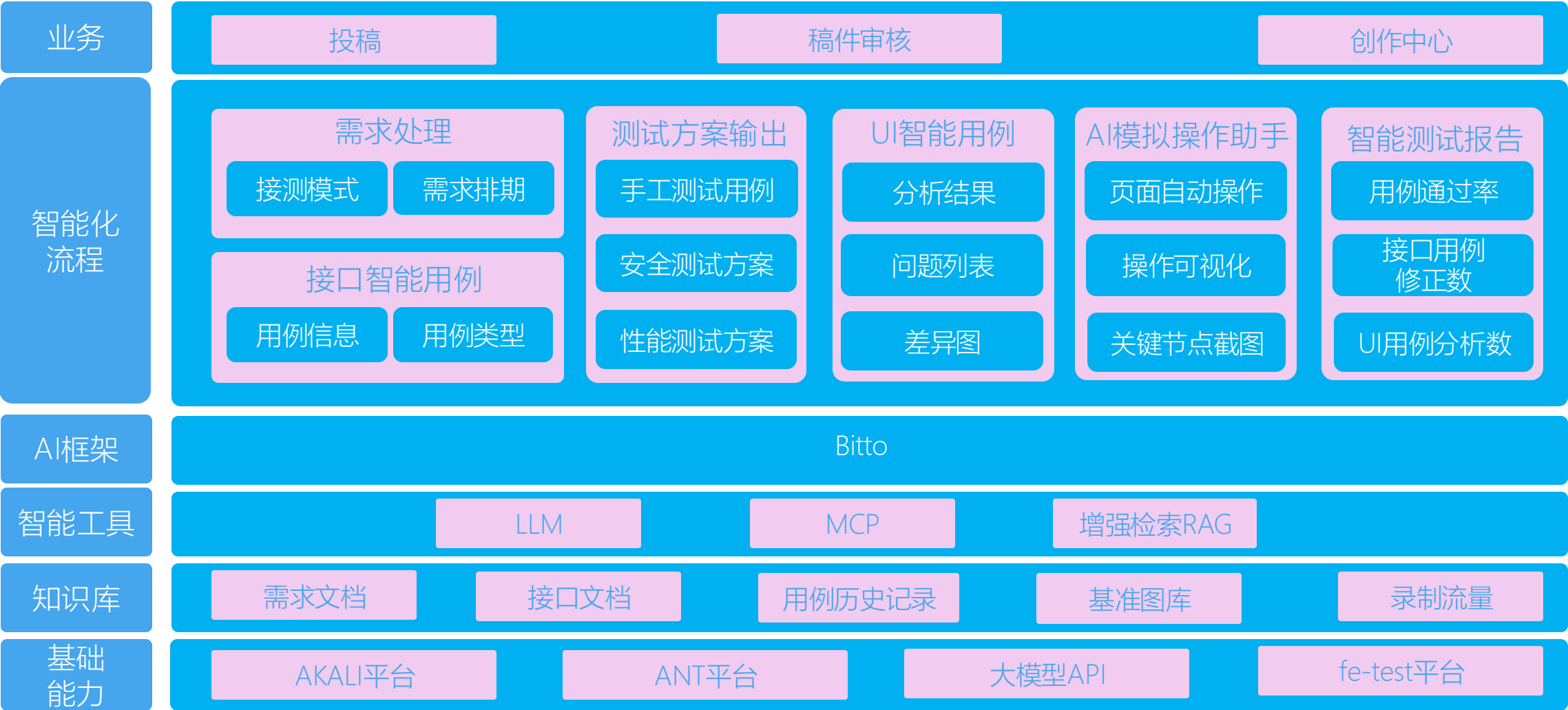
解决方案

解决思路

整体思路：利用AI的学习和推理能力完成常态化的测试活动



整体架构



整体设计

生产流程

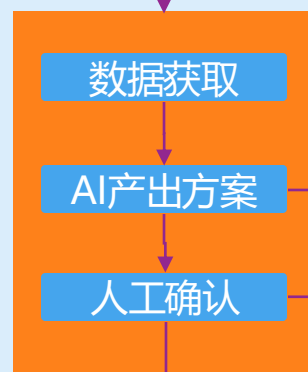


需求分级

接口智能用例生成

UI智能用例测试|智能测试报告

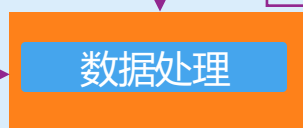
Agent



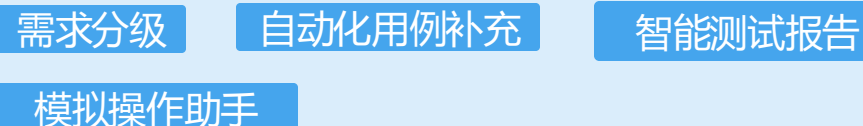
获取方案

知识学习

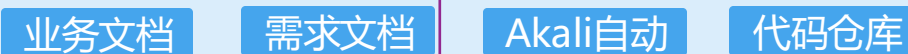
结果反驳



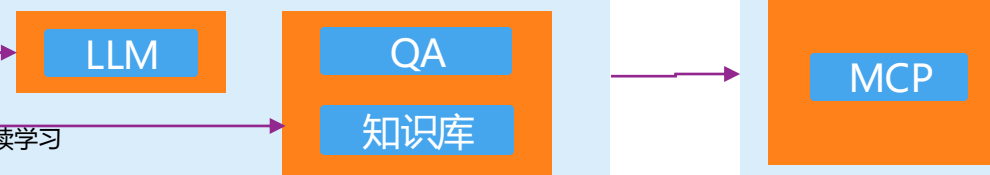
产出



数据



AI



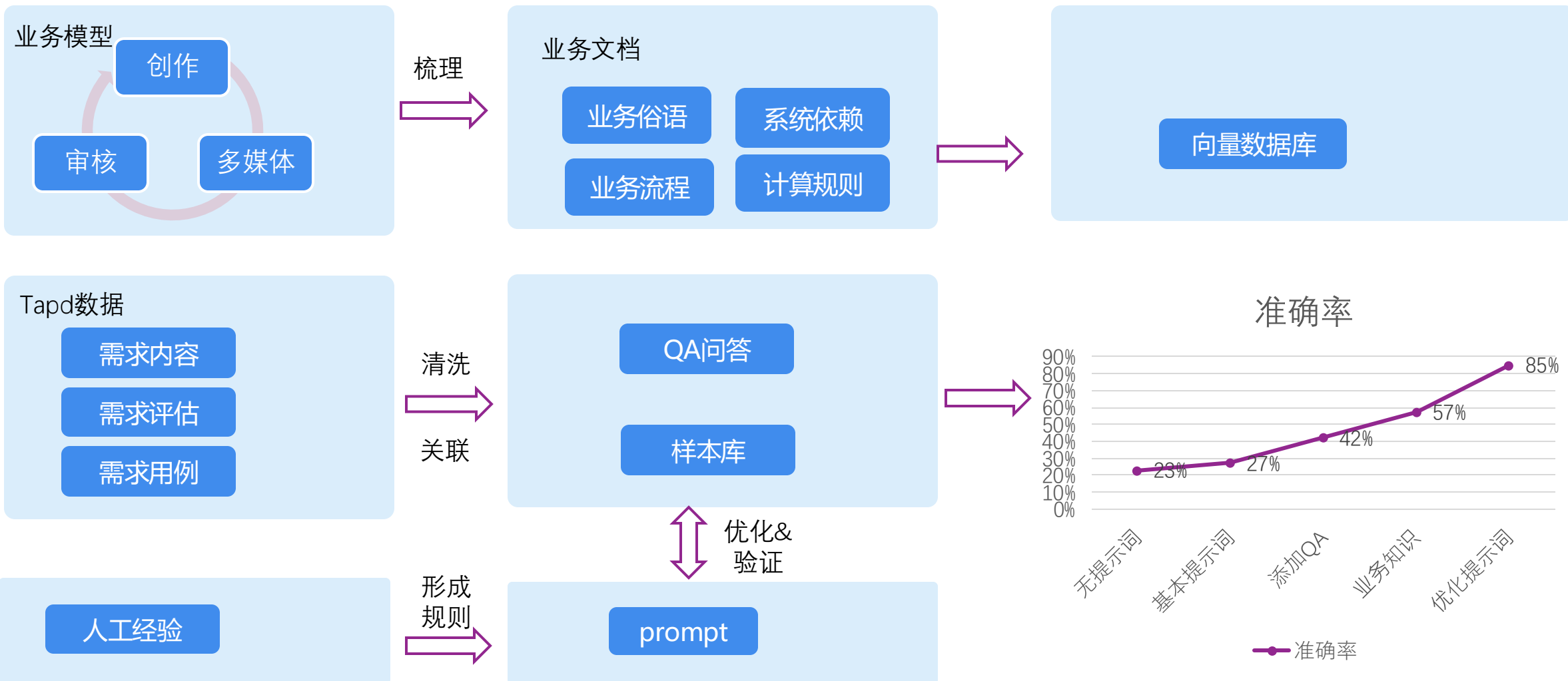
持续学习

PART 04

具体实现

需求分级实践-数据准备

阶段



需求分级实践-实践结果

Step1: 流程触发

基本信息

状态

需求评审通过

分类

工作台

需求来源

产品规划

需求方

--

服务OKR目标

工作台 (Tag)

Step2: AI分级结果

需求分级 测试用例

AI风险分析:

项目评级为A级, 接测分数直接置为1, 需求内容复杂度评分为0.8, 技术文档复杂度评分为0.5, 开发工时为4d以上, 舆情等级为0, 综合以上数据得出接测分数为0.8, 属于正式接测范围。

AI处理建议:

需求风险

0.00

0.80

1.00

接测方式

正式接测

Step3: 手工用例结果

	用例编号	步骤	预期
0	TC-AUTO-01	提交符合活动规则的活动稿件 (时长、原创、标签、来源均合规)	活动平台自动判断通过, 无需依赖审核平台
1	TC-AUTO-02	提交活动稿件时设置时长不符合活动规则	活动平台自动判断拒绝并返回具体时长错误提示
2	TC-AUTO-03	提交非原创内容但标记为原创的活动稿件	活动平台自动判断拒绝并返回原创性校验失败提示
3	TC-AUTO-04	提交包含禁用标签的活动稿件	活动平台自动判断拒绝并返回标签违规提示
4	TC-AUTO-05	提交来源渠道不在白名单内的活动稿件	活动平台自动判断拒绝并返回来源不合法提示
5	TC-MANUAL-01	人工审核符合画风规范的活动稿件	审核平台显示通过标识, 稿件进入下一流程
6	TC-MANUAL-02	人工审核不符合画风规范的活动稿件	审核平台显示拒绝标识并允许填写驳回原因
7	TC-AUTO-06	在活动平台修改判断规则后提交历史驳回稿件	根据新规则重新判断结果, 原驳回原因失效时自

导出为csv

导出为xmind

导出为pdf

Step4: 更新结果

免测 ? 否

测试人员

预计开始

--

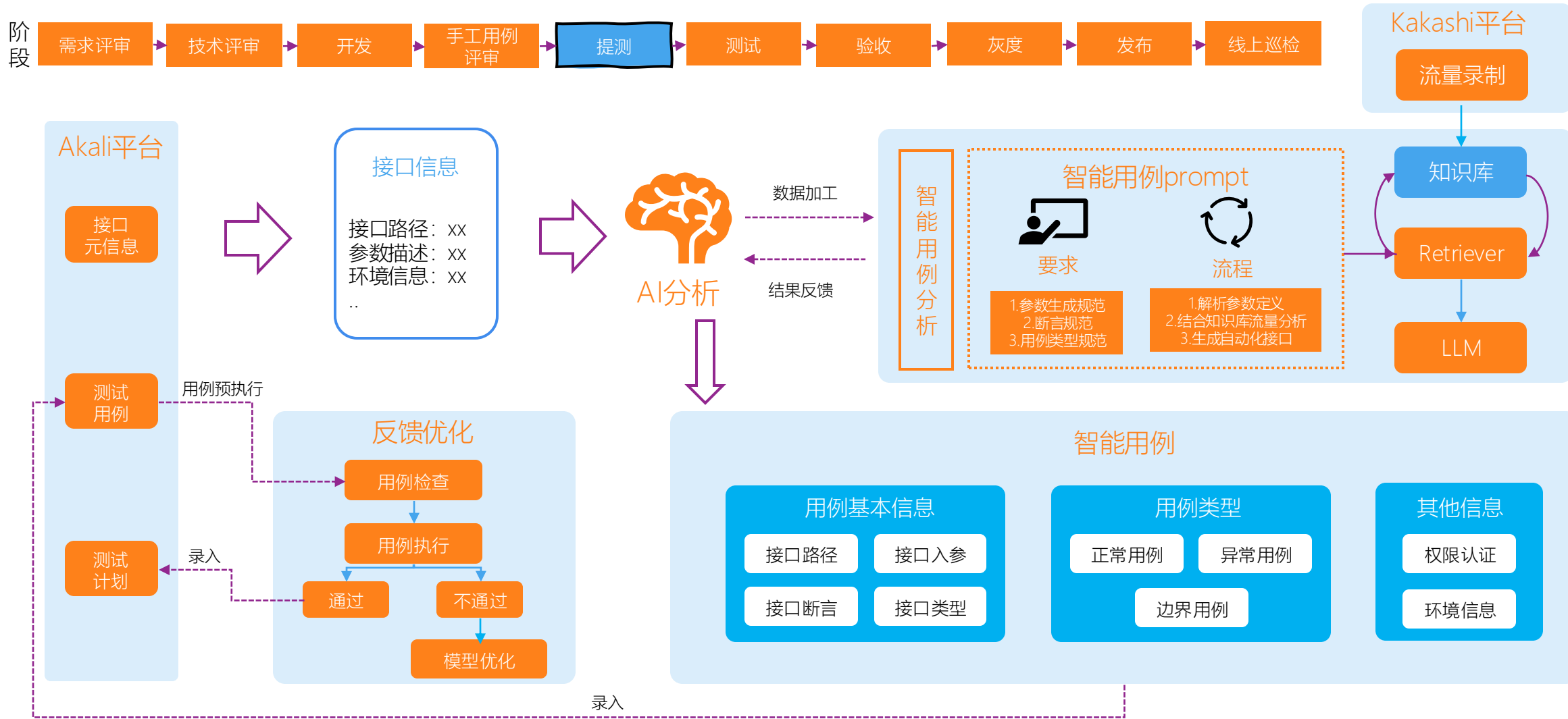
预计结束

--

预计联调时间

--

具体实现—接口智能用例



应用实践—接口智能用例

Step1: 选择流量集



Step3: 根据接口信息生成用例



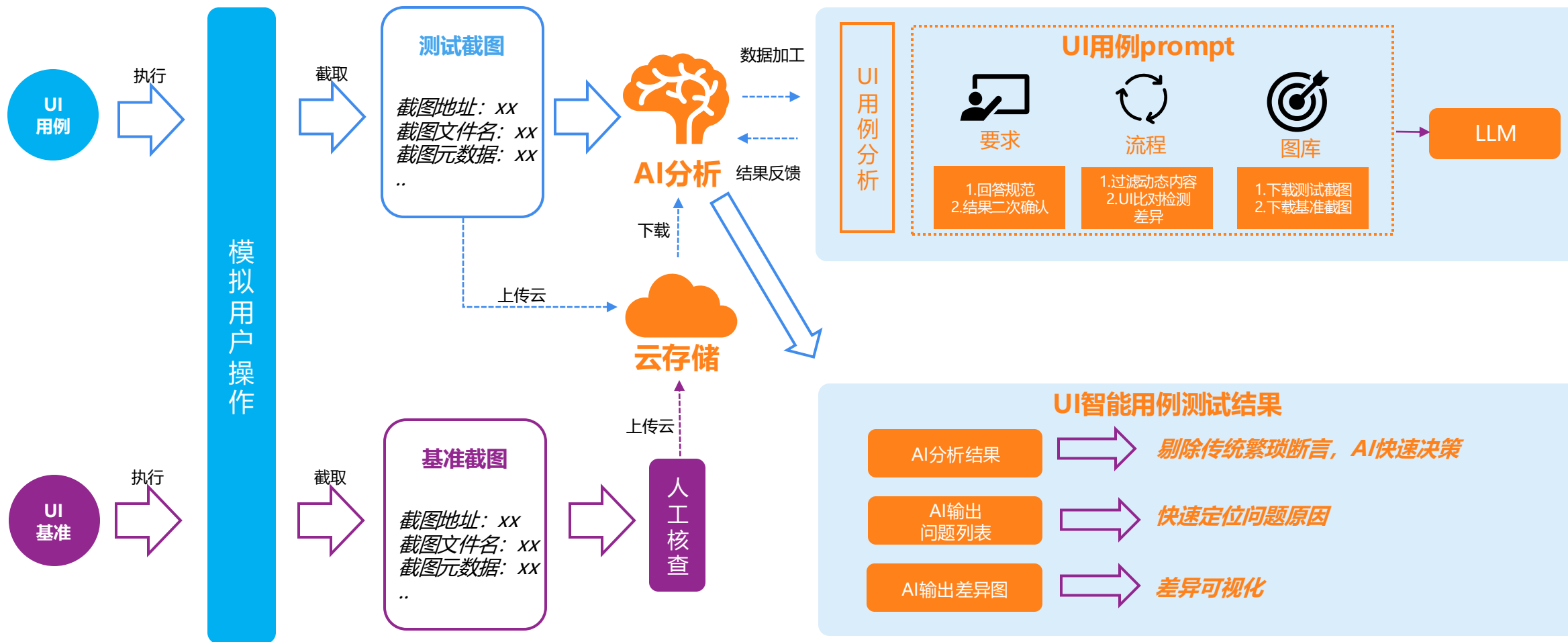
Step2: 解析流量集



Step4: 生成测试报告



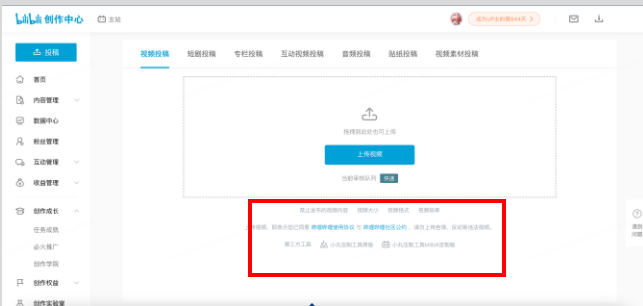
具体实现—UI智能用例



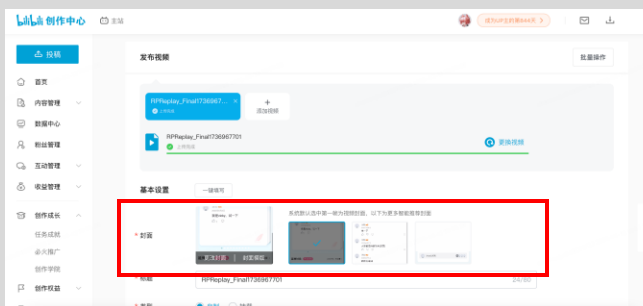
应用实践—UI智能用例

UI智能用例案例

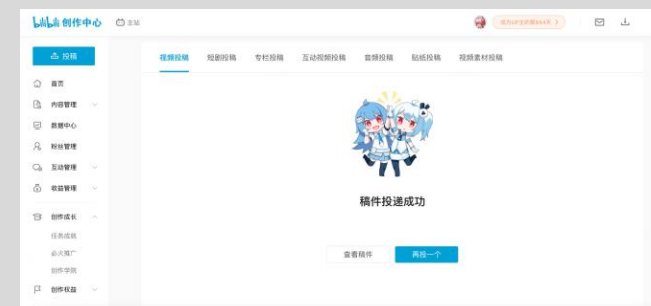
基准截图



对比(缺少文案)

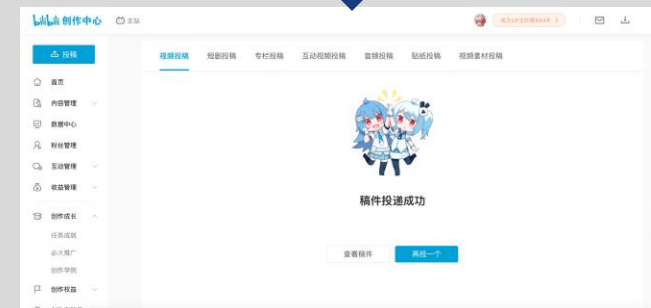
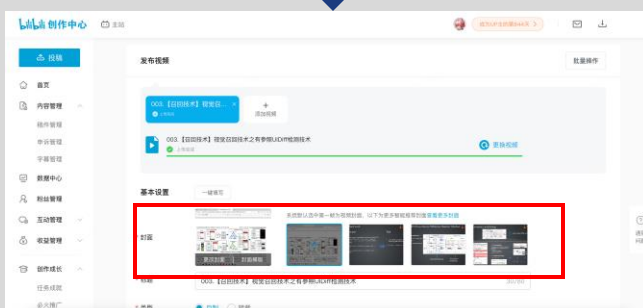
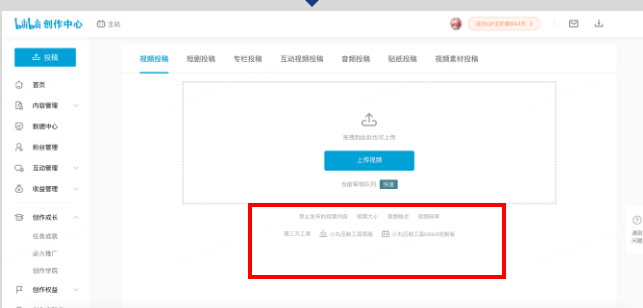


对比(动态区域不一致)

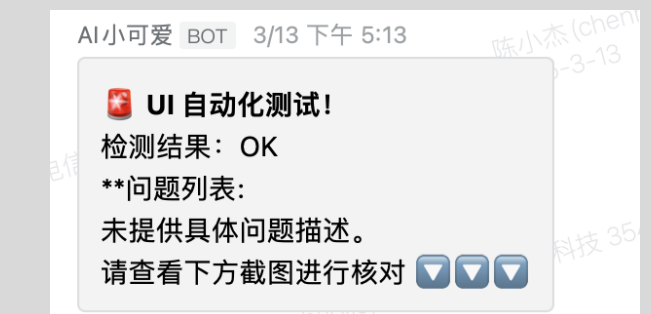
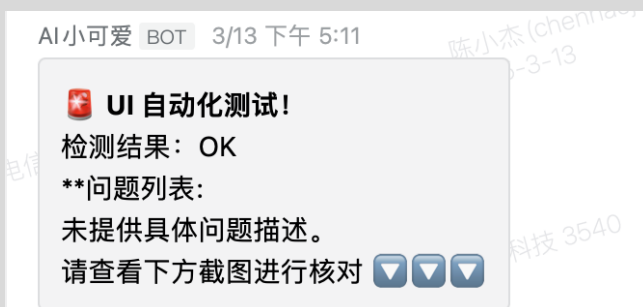
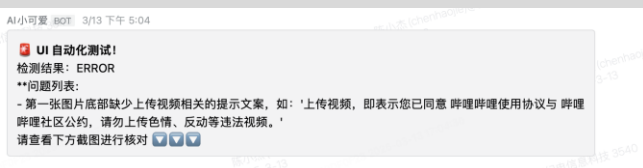


对比(完全一致)

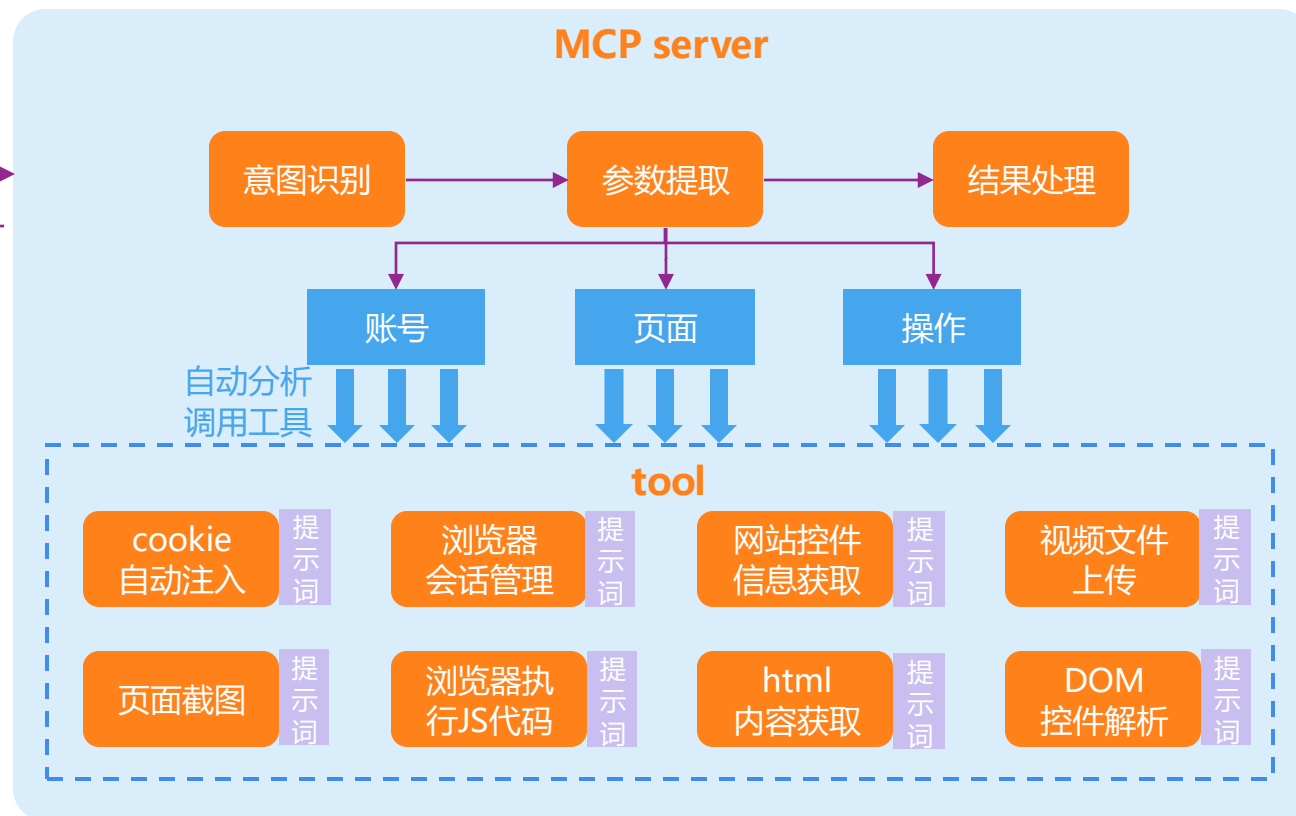
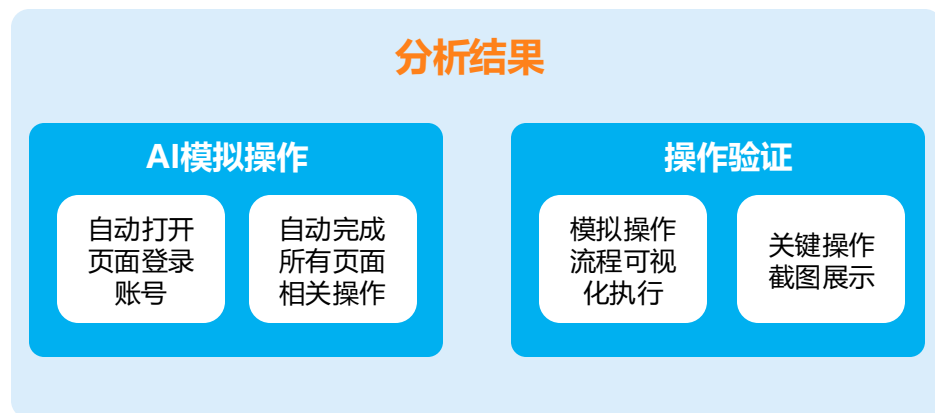
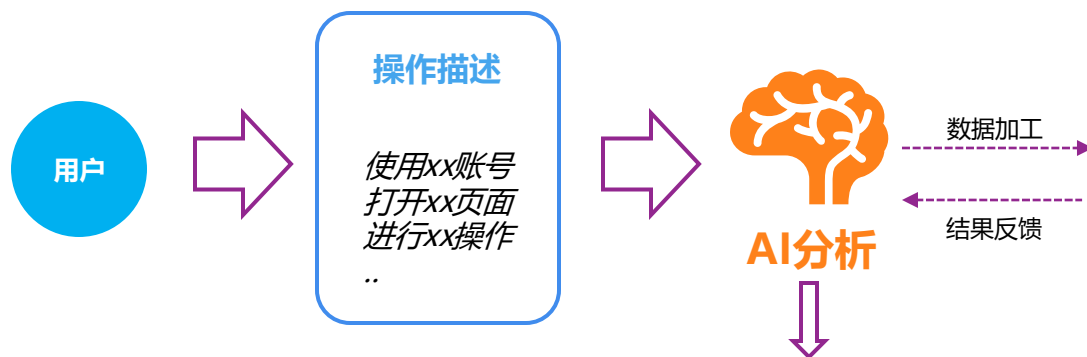
本次测试



对比结果



具体实现—AI模拟操作助手



应用实践—AI模拟操作助手

AI模拟操作案例

Step1: 用户提问



用户

05/07 11:22

使用playwright-bili-test,用mid=1700000000000000测试投稿页面,你要执行以下步骤

1. 打开投稿页面, <https://member.bilibili.com/platform/upload/video/frame>
2. 帮我上传视频 /Users/futangyi/Documents/测试视频\ 2.mov
3. 修改标题为ty212
4. 分区修改为影视
5. 点击立即投稿

你需要获取投稿页的常用控件, 操作方法优先从常用控件中获取

Tokens: 164

Step2: claude 模型分析

我将按照您的要求, 使用playwright-bili-test来测试B站的投稿页面。首先, 我需要获取cookie, 然后创建浏览器会话, 并执行您列出的步骤。

首先, 我需要获取Bilibili的cookie:

现在我已经获取到了cookie, 接下来我将创建一个浏览器会话并导航到投稿页面:

现在我已经成功创建了浏览器会话, 接下来我需要获取投稿页面的常用控件信息:

现在我已经获取了投稿页面的常用控件信息, 接下来我将按照步骤执行操作:

1. 首先, 我需要上传视频文件:

文件已成功上传。现在我需要等待一下, 确保视频上传完成, 然后进行截图查看当前状态:

Step3: mcp tool 执行

bilibili_get_cookie_by_mid | 已完成 ✓

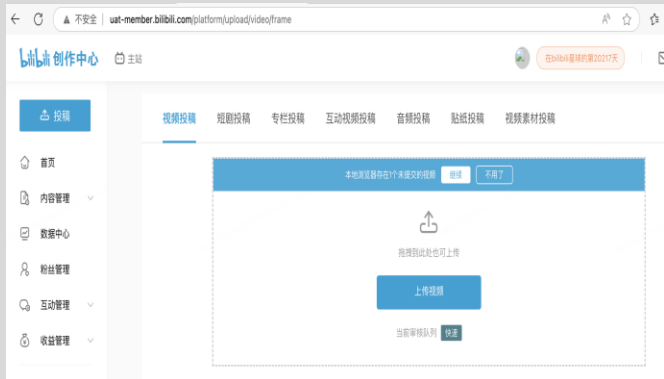
playwright_new_session | 已完成 ✓

get_controls | 已完成 ✓

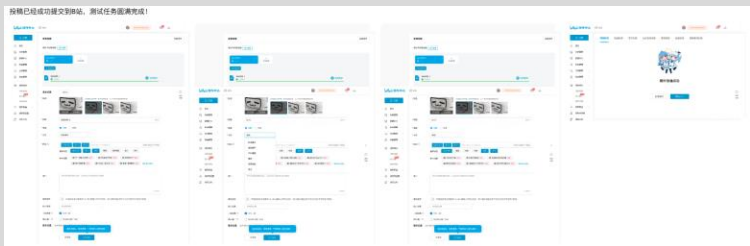
playwright_upload_file | 已完成 ✓

playwright_screenshot | 已完成 ✓

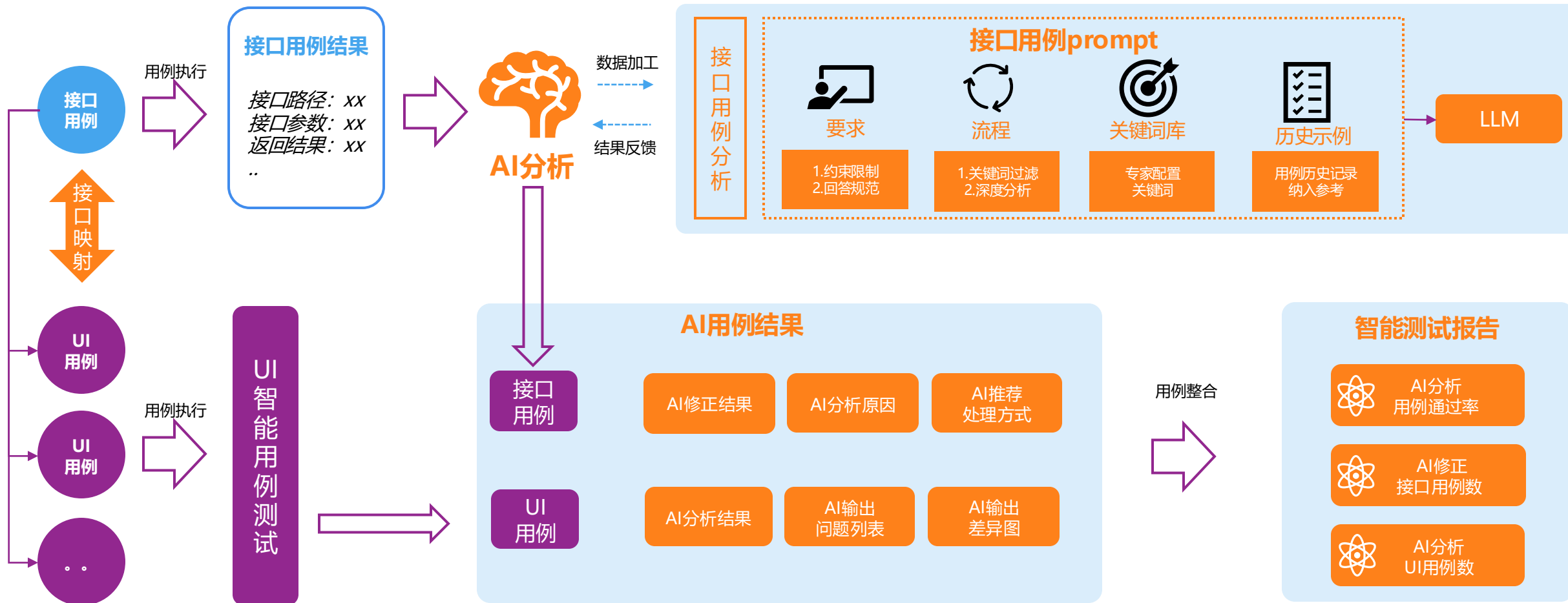
Step4: 模拟操作 执行



Step5: 截图查看 验证



具体实现—智能测试报告



应用实践—智能测试报告

接口自动化用例执行不稳定案例

用例: /x2/creative/app/season/switch

>	成功	GET /x2/creative/app/seasons	/x2/creative/app/seasons 查 状态码: 200 耗时: 31
>	成功	POST /x2/creative/app/season/section/episodes/add	/x2/creative/app/season/section/episode 添加单集
√	失败	POST /x2/creative/app/season/switch	/x2/creative/app/season/switch 状态码: 200 耗时: 5

开始时间: 2025-05-07 16:38:49 结束时间: 2025-05-07 16:38:49

响应header 响应body 断言 提取 请求内容

1 [{"code": -400, "message": "invalid character ' ', looking for beginning of value", "ttl": 1}] [{"code": -500, "message": "服务器错误", "ttl": 1}]

案例描述: 用例节点执行出现服务器错误, 并非代码变更导致的错误

Ai分析结果

用例: /x2/creative/app/season/switch

✨ AI修正 命中“服务器错误”

该用例经过AI分析发现命中关键词(服务器错误), AI修正结果为成功, 符合预期

用例: /x2/internal/creative/season/creative/auth

>	成功	GET /x2/internal/creative/season/creative/auth	合集创作无权限 状态码: 200 耗时: 19
√	失败	GET /x2/internal/creative/season/creative/auth	合集创作有权限 状态码: 200 耗时: 161

开始时间: 2025-05-15 16:33:19 结束时间: 2025-05-15 16:33:19

响应header 响应body 断言 提取 请求内容

断言名称	断言内容	错误信息	断言脚本	是否成功
Assertion	\$.data expect: true	Value equals to be 'true', but found 'false'		失败

案例描述: 用例节点出现断言报错, 确实是用例存在问题导致的错误

Ai分析结果

用例: /x2/internal/creative/season/creative/auth

✨ AI分析 | 失败原因 接口返回的 data 字段值为 false, 但预期值为 true, 表示用户没有合集创作权限。可能因为用户权限配置错误或后端逻辑异常。”

AI快速分析出失败原因, 判断出用例执行本身存在问题, AI判定结果为失败, 符合预期

Prompt

LLM

Result

AI流程典型问题

AI流程	AI模拟操作助手+UI智能用例	接口智能用例	智能测试报告	AI智能分级
问题 & 解决方式	1. css组件信息人为补充困难，扩展性差 底层使用mcp+claude，所有组件信息支持自动分析调用 2. 难以确定模拟操作流程是否准确 mcp加入截图工具 截图工具加入特定提示词 3. AI容易对以下UI差异产生误判 - 微小像素偏移 / 字体渲染差异 / 阴影 / 背景渐变 - 页面中的动态内容（如时间戳、广告、头像） 提示词工程优化： 只关注“明显 UI 异常”：如元素缺失、错位、颜色严重错误，忽略“微小差异”和“动态内容”	1. AI获取接口文档数据，数据中容易混杂 Markdown、表格、自然语言描述 限定生成结构 引导模型提取核心字段 2. AI推理接口路径、字段命名、示例值不标准 知识库增加历史用例	1. 用例执行失败存在明显差异，大部分是用例抖动导致，极少部分是代码原因导致，AI难以作出准确判断 提示词加入第一层关键词过滤 提示词加入第二层深度分析 知识库增加历史用例执行结果	1. 业务特性差异 针对不同业务，采用配置产生提示词。例如阈值、业务核心模块等 2. 针对无效内容生成用例 需求内容清洗，去掉背景收益等。 3. 用例采纳率较低 历史逻辑需要知识库形式存在 提示词抢用例提示提供参考依据，否则不生成
收益	测试结果截图直观展示，功能可快速覆盖各个主流程，UI智能用例测试结果可信度增加	接口智能用例生成格式规范，用例生成准确性明显提升	关键词过滤可准确解决95%以上的失败问题，剩余5%通过深度分析可给出有效建议和AI修正结果	排除需求内容中无关部分，能有效减少干扰，增加准确性和有效率



1

需求处理

- **需求分级**：单次需求沟通会，接测评估效率提升2h

2

测试阶段

- **手工用例生成**：单次需求手工智能用例占比达到40%以上
- **AI模拟操作助手**：降低测试人员编写自动化用例的成本，可通过描述性语言智能驱动UI自动化测试，AI模拟操作助手使用率占全组80%以上，累计完成2000次以上模拟操作，召回17个有效问题
- **UI智能用例**：高频迭代页面用例复用率由20%提升至90%，维护成本下降近60%，单人每周节省维护工时约3h
- **接口智能用例**：平均每个复杂接口提升用例设计与编写时间约20min

3

用例维护

- **智能测试报告**：1h/天的人力用例问题排查投入降低到0.2h/天的人力用例问题排查



PART 05

总结与展望

AI赋能的智能化测试体系实现了全链路质量保障升级，极大的降低了人力投入，提高了测试效率；该流程目前还在起步阶段，距离完全实现全流程智能化还有很长的路要走，需要大家继续努力。



- 应用AI能力时进一步控制资损，提升性能及稳定性
- 构建测试知识持续学习系统，实现模型能力的动态进化
- 未来的接口智能用例应从单一接口验证向多阶段场景智能化演进
- UI智能用例测试和AI模拟操作能力的愈加完善将完全替代前端页面的测试工作



参与调研您将优先获得



AiDD定制版
《AI+软件研发精选案例》



专属学习顾问
1对1需求对接

AiDD会后小调研

AiDD峰会致力于协助企业利用AI技术深化计算机对现实世界的理解,推动研发进入智能化和数字化的新时代。作为峰会的重要共建者,您的真知灼见对我们至关重要。衷心感谢您的参与支持!



扫码参与调研

2025 AI+研发数字峰会

拥抱 AI 重塑研发

科技生态圈峰会 + 深度研习

——1000+ 技术团队的选择



敦煌站

K+ 思考周®研习社

时间: 2025.08.29-30



上海站

K+ 金融专场

时间: 2025.09.26-27



香港站

K+ 思考周®研习社

时间: 2025.11.17-18



K+峰会详情



上海站

AI+研发数字峰会

时间: 2025.05.23-24



北京站

AI+研发数字峰会

时间: 2025.08.08-09



深圳站

AI+研发数字峰会

时间: 2025.11.14-15



AiDD峰会详情

AiDD 峰会



2025 AI+研发数字峰会
AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

