

大模型测试域运用落地实践介绍

技术中心测试部 郝晓升

2023年11月2日

01 背景分析

02 AITEST功能介绍

03 AITEST与提词平台结合

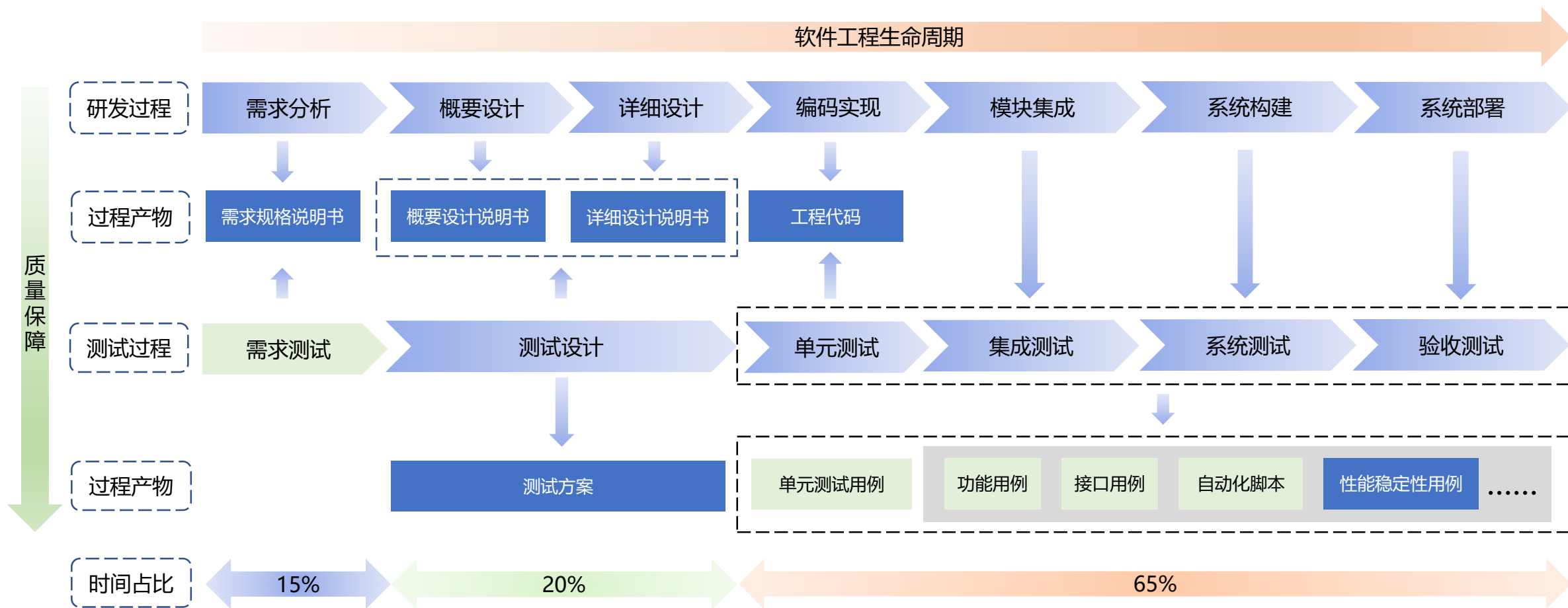
04 如何提升大模型的生成效果

05 未来展望

背景分析

背景分析

大模型技术浪潮下，将大模型融入到软件工程生命周期中的各个环节，实现研测提效



需求分析及研发测试过程约占软件工程生命周期80%时间，可将大模型应用到两个阶段，提升过程产物输出的效率及质量，实现整体提效

背景分析

典型案例一	典型案例二	典型案例三	典型案例四	典型案例五
实践场景	实践场景	实践场景	实践场景	实践场景
<ul style="list-style-type: none">需求测试	<ul style="list-style-type: none">功能测试用例设计	<ul style="list-style-type: none">接口测试脚本生成	<ul style="list-style-type: none">自动化测试脚本生成	<ul style="list-style-type: none">测试工具脚本生成
业务线	业务线	业务线	业务线	业务线
<ul style="list-style-type: none">教育课堂应用中心	<ul style="list-style-type: none">教育课堂应用中心	<ul style="list-style-type: none">民族共同体业务线	<ul style="list-style-type: none">教育课堂应用中心	<ul style="list-style-type: none">教育学习机
验证价值点	验证价值点	验证价值点	验证价值点	验证价值点
<ul style="list-style-type: none">大模型提供15条需求疑问，结合业务上下文有价值疑问7条，采用率47%。可以在需求阶段直接给产品经理生成疑问帮助产品优化需求	<ul style="list-style-type: none">大模型返回用例设计39条，无效用例5条，准确率34/39=87%。生成的用例还需要优化去补充其它场景	<ul style="list-style-type: none">生成10条用例，可以直接执行成功5条，有5条需要人工介入调试，可以快速生成接口测试用例脚本，生成后只需要人工进行简单调试就可以运行。脚本效率提升50%	<ul style="list-style-type: none">大模型实现大部分用例从步骤描述到测试脚本的转化，转化后还需要人工介入调试。单个用例脚本编写耗时可以由2小时缩短到1小时，效率提升50%	<ul style="list-style-type: none">大模型可以实现大部分当前测试过程中的大部分小工具脚本生成，效率提升50%+
问题点	问题点	问题点	问题点	问题点
<ul style="list-style-type: none">由于token的限制，无法一次输出完整的项目需求，只能给大模型输入部分需求	<ul style="list-style-type: none">与需求测试问题相同	<ul style="list-style-type: none">部分用例需要人工介入调试	<ul style="list-style-type: none">需要人工介入调试	<ul style="list-style-type: none">需求描述要清楚，生成的脚本无法保证100%可以直接运行需要做好调试的准备

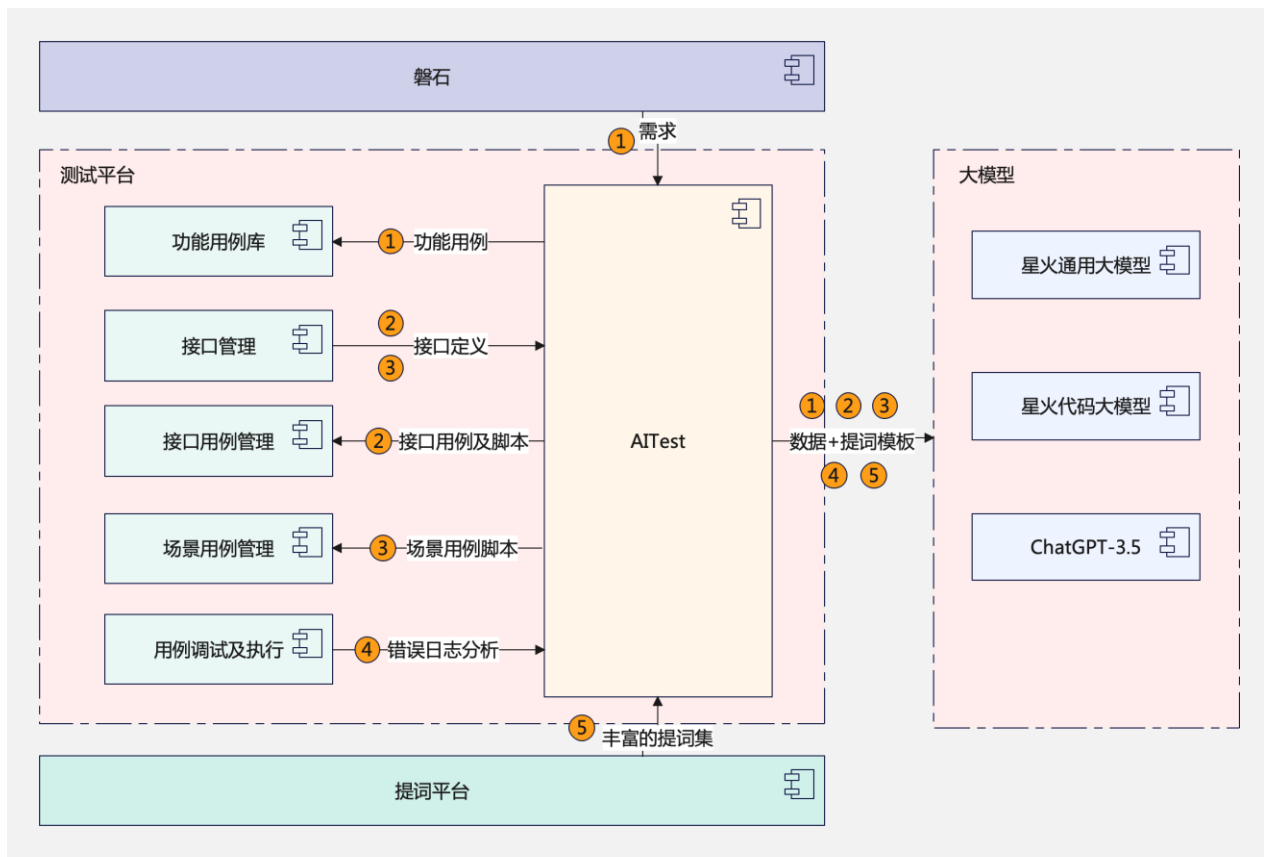
实践发现，大模型可在需求测试、用例设计及自动化测试等方面发挥较大作用，辅助测试人员提升效率

AITEST功能介绍

AITEST功能介绍

功能概述

通过广泛的大模型测试实践，对最佳实践案例进行了平台化能力封装，实现了测试域提词的集中管理及个性化修改定制，打通并整合了多个平台能力，将AI能力与工具平台相结合，持续提升测试活动各个环节的效率



- 1 从磐石获取需求，可分钟级生成符合集团统一用例编写规范的功能测试用例，可便捷导入测试平台
- 2 从接口管理模块获取单接口定义，可分钟级生成单接口测试用例及测试脚本，接口用例可导入到测试平台进行管理维护
- 3 从接口管理模块获取多个接口定义，生成场景用例脚本，可导入测试平台进行管理维护
- 4 对执行报错的用例进行日志分析，给出可能的原因，辅助测试人员定位问题
- 5 对接提词平台更多测试能力，可对提词进行自定义修改保存，平台具备推荐优秀提词的能力，实现提词的共享共建

AITEST功能介绍——测试设计领域

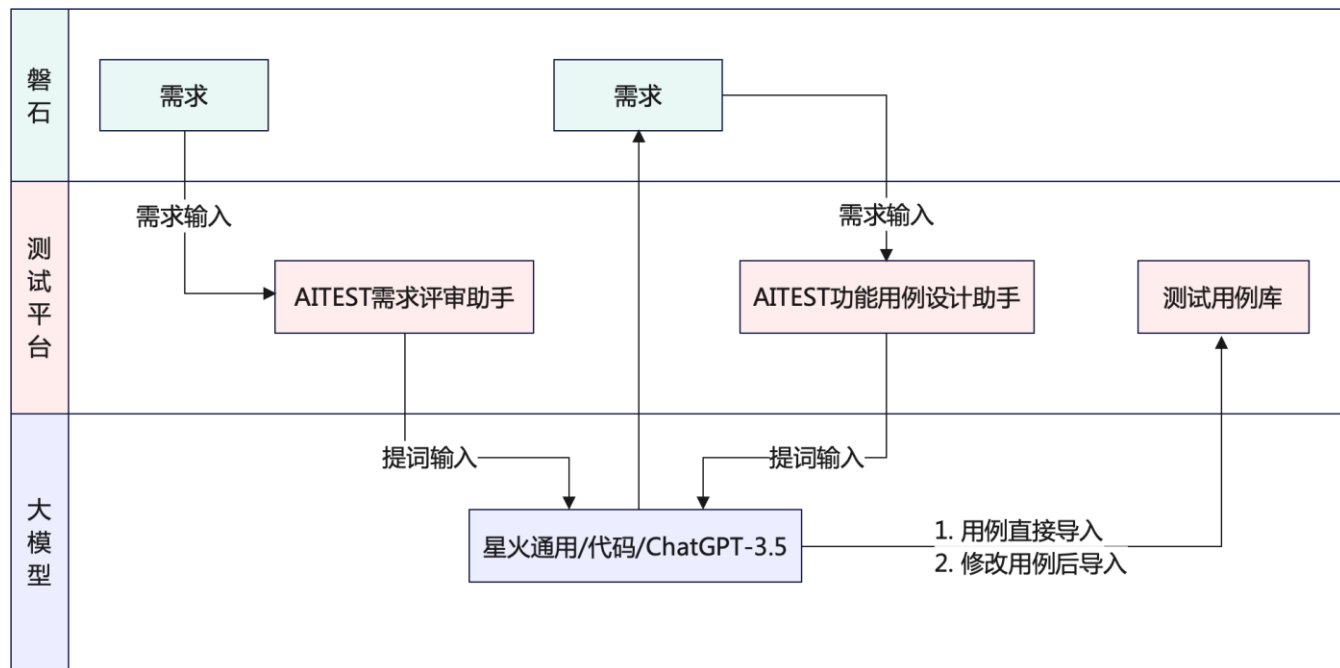
大模型在测试设计领域中的应用

● 需求测试

- 针对需求编写不规范，需求描述不清晰的问题，可通过大模型对需求进行评审，发现需求说明中存在的问题，在需求迭代频繁的情况下，可以提升需求评审的效率

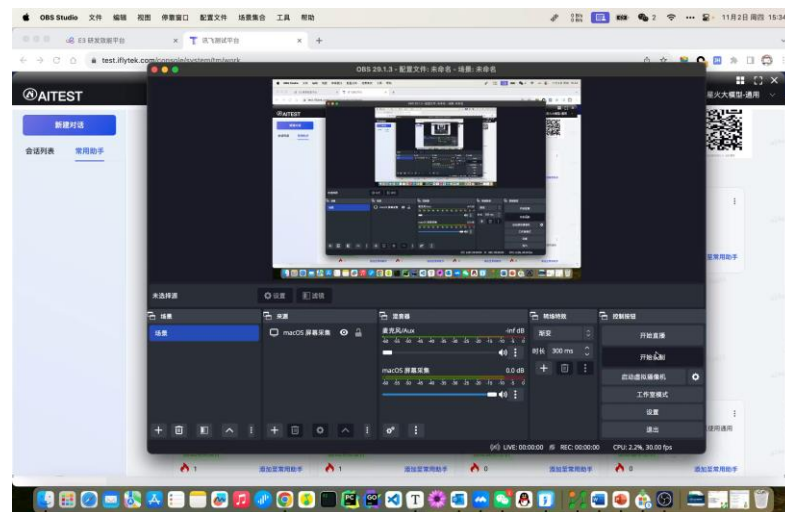
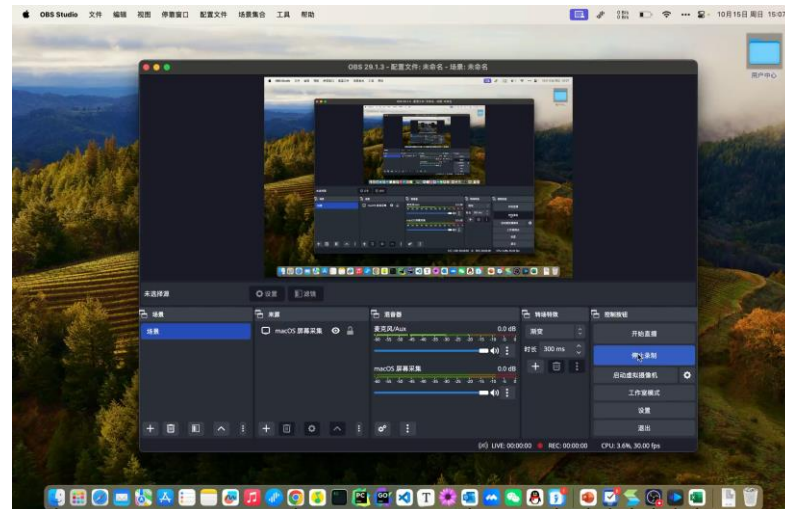
● 功能用例设计

- 不同的测试人员及业务线，用例设计规范不同，频繁的产品演进用例设计耗时，复杂的业务系统功能模块众多，人工编写用例较为耗时，使用大模型可以根据需求快速生成符合规范的测试用例，测试人员可在此基础上进一步完善，提升用例编写效率



大模型的引入，提升了需求评审及功能用例设计的效率

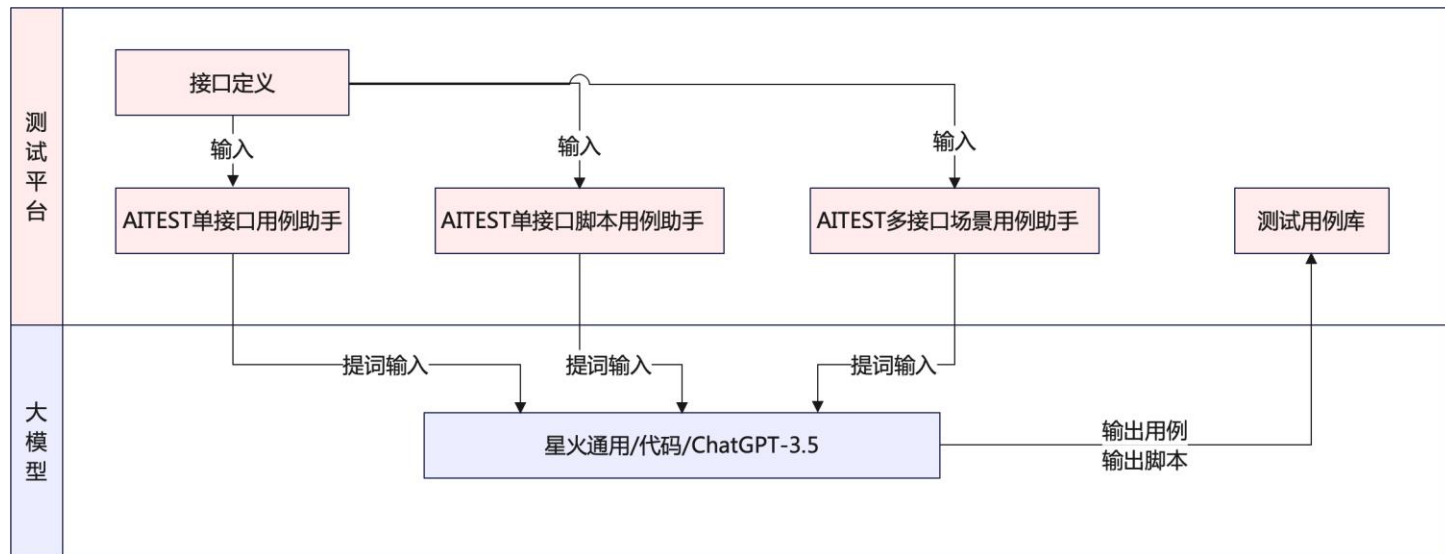
平台功能演示



AI TEST功能介绍——自动化测试领域（接口）

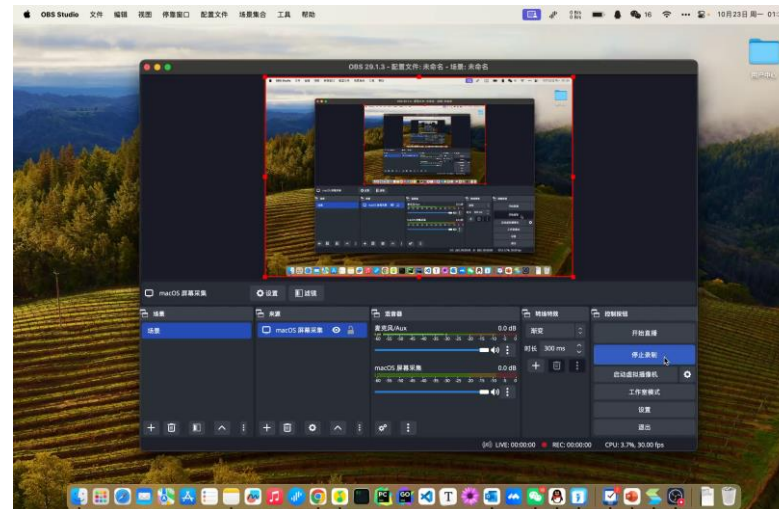
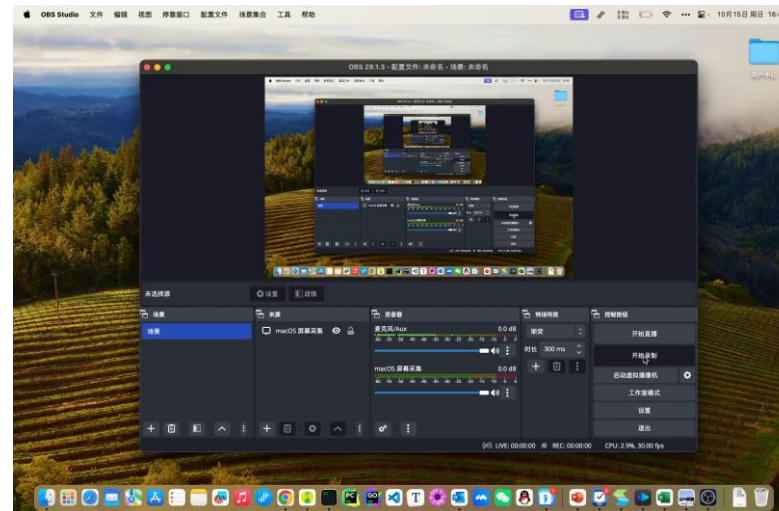
大模型在自动化测试领域的应用

- 接口自动化用例及脚本生成
 - 面向所有测试人员，提供接口测试用例及脚本自动生成的能力，根据接口定义文档输出测试用例及脚本，降低脚本编写的门槛，提高接口测试用例及脚本编写的效率
- 场景自动化用例及脚本生成
 - 根据多个接口定义文档，自动编排生成场景级测试脚本，提升场景测试脚本编写效率



大模型的引入，减少了脚本编写的时间，但输出的代码还需要人工调试

平台功能演示



AITEST功能介绍——自定义提词



自定义提词

- ❑ 可添加个人助手
- ❑ 可借鉴查看其它优秀助手
- ❑ temperature的参数值越小，模型就会返回越确定的一个结果。如果调高该参数值，大模型可能会返回更随机的结果，也就是说这可能会带来更多样化或更具创造性的产出

temperature number or null Optional Defaults to 1

What sampling temperature to use, between 0 and 2. Higher values like 0.8 will make the output more random, while lower values like 0.2 will make it more focused and deterministic.

We generally recommend altering this or `top_p` but not both.

top_p number or null Optional Defaults to 1

An alternative to sampling with temperature, called nucleus sampling, where the model considers the results of the tokens with top_p probability mass. So 0.1 means only the tokens comprising the top 10% probability mass are considered.

We generally recommend altering this or `temperature` but not both.

AITEST与提词平台结合

AI TEST与提词平台结合

AI TEST集成提词平台15种测试域提词能力，可对提词进行自定义修改，保存为个人助手，方便后续使用

提词平台

方案设计

用例设计

脚本开发

测试结果分析

性能测试

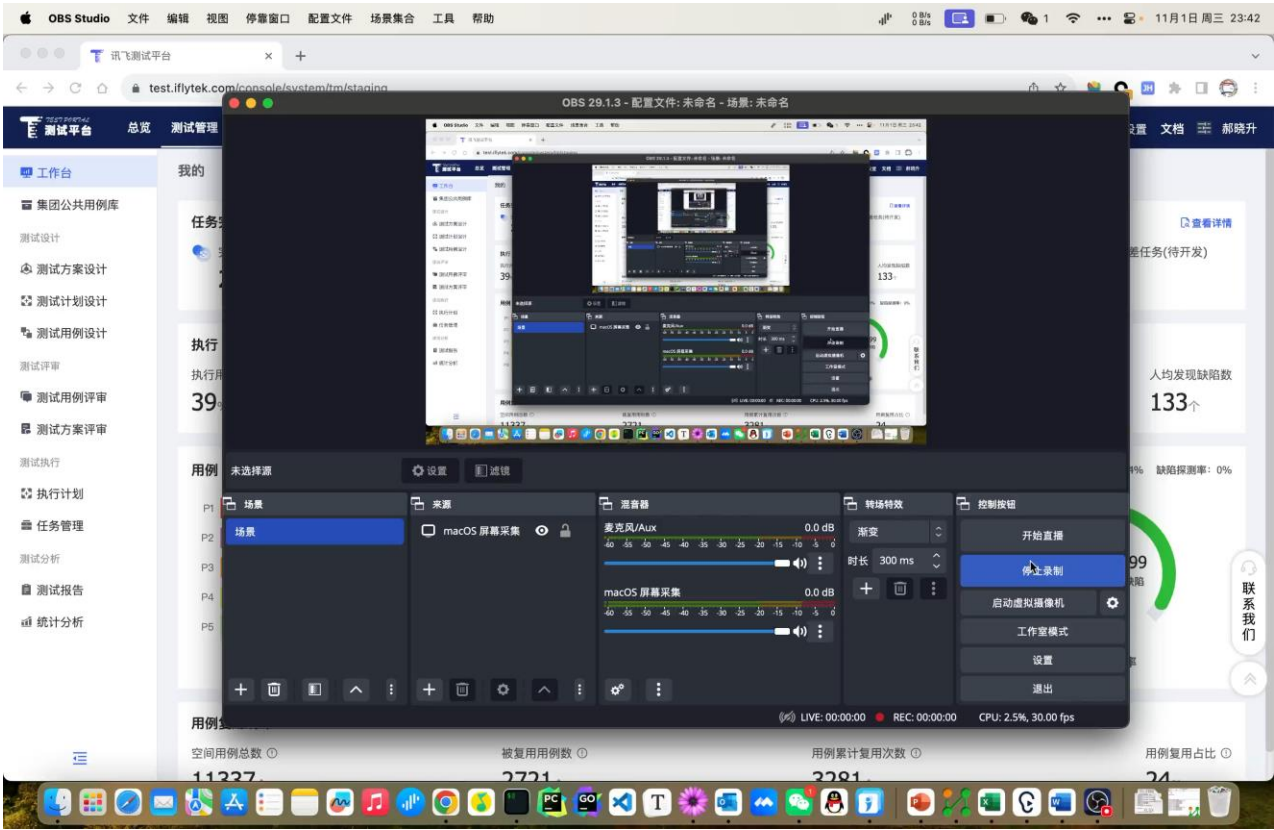
测试工具开发

测试数据准备

对比测评

测试学习

.....



如何提升大模型的生成效果

如何提升大模型的生成效果

当前问题

大模型生成结果的质量强依赖提词输入的质量，由于输入需求或输入接口不规范、不完整，导致输出的功能用例或接口用例不完善

从输入数据及提词模型两方面入手，提升大模型生成效果



解决路径

- ❑ 对于输入的需求，避免大篇幅的输入，提炼需求的功能点，以小粒度的功能点输入给大模型，提升功能用例的准确性和覆盖度
- ❑ 建立需求知识库，完善需求背景知识，使输入需求更饱满更全面
- ❑ 推进接口定义规范；针对不同的接口参数，设计不同的用例生成规则；增加现网导流数据，辅助大模型生成接口调用关系

如何提升大模型的生成效果

示例提示词

你是一个经验丰富的高级系统测试工程师，现在根据项目需求设计功能测试用例。如果你不了解这些需求，可以访问bing搜索完善产品和背景信息。如果你判断需求不完善，可以先完善需求。慢慢思考，不用着急答复。测试用例设计的要求如下：

~~~

- 1、以markdown语法、表格的格式输出用例。
- 2、用例表格包含字段：用例名称、优先级、测试项、前置条件、步骤详情、预期结果、测试类型、是否反向用例，并显示在第一行。
- 3、用例名称按照主谓宾格式表达。
- 4、字段优先级定义如下：P1：核心功能的正向测试用例（冒烟）；P2：基本功能的负面测试用例；P3：核心功能反向测试用例(包含重要的错误和边界值)；P4：基本功的反向测试用例（异常测试，边界）；P5：不常用功能的测试用例（不经常被执行）。
- 5、测试项用于概括用例的被测对象，需简洁，5个中文汉字以内。
- 6、覆盖所有符合有真实用户场景，并覆盖所有的功能点。
- 7、尽可能详尽地覆盖用户异常场景，确保浏览和提交无异常。
- 8、设计的用例需要覆盖第9步中说明的全部测试类型。
- 9、测试类型可选项为"功能测试、兼容性测试、易用性测试、性能测试、稳定性测试、可靠性测试、接口测试、安全测试"。

~~~

需求如下：

{{需求}}

提示词要素

角色设定

角色提示技术是通过为大模型指定一个特定的角色来引导其输出的一种方式。这种技术对于生成针对特定上下文或受众的文本非常有用

明确指令

指令提示技术是通过为模型提供具体指令来引导大模型输出的一种方法。这种技术对于确保输出相关和高质量非常有用

上下文

包含外部信息或额外的上下文信息，引导大模型更好地响应

鼓励式引导

鼓励式引导是一种技巧，可鼓励大模型生成反思和思考性的文本

输入数据

用户输入的内容或问题

输出提示

指定输出的类型或格式

➤ <https://www.promptingguide.ai>

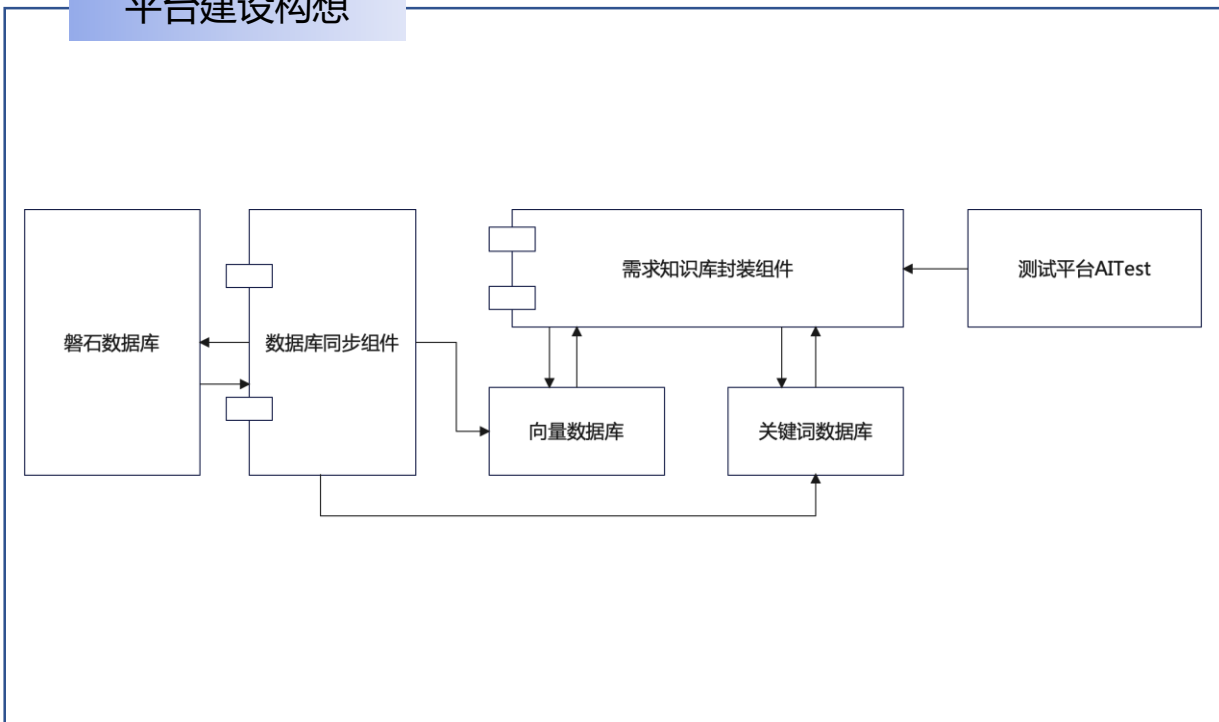
➤ <https://github.com/ORDINAND/The-Art-of-Asking-ChatGPT-for-High-Quality-Answers-A-complete-Guide-to-Prompt-Engineering-Technique>

未来展望

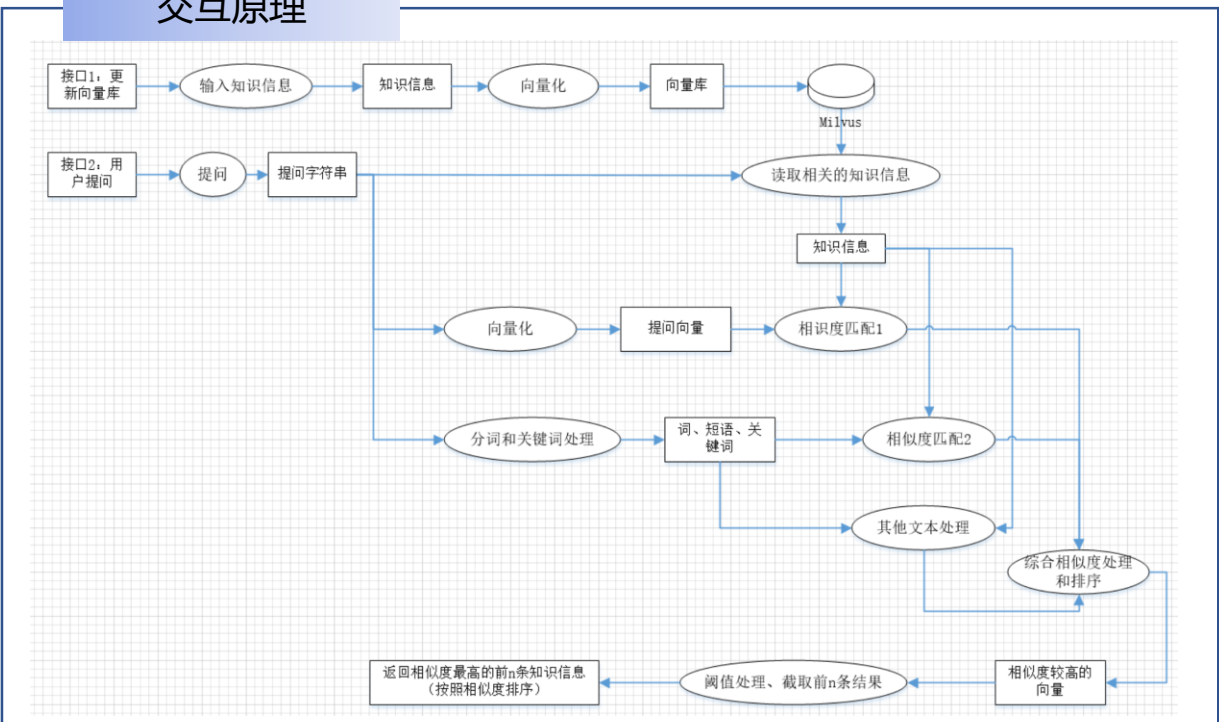
未来展望——提升用例设计效果

建立需求知识库，通过检索待生成用例的需求关联的需求集，提升需求描述的完善度，从而提高功能用例的完善度

平台建设构想

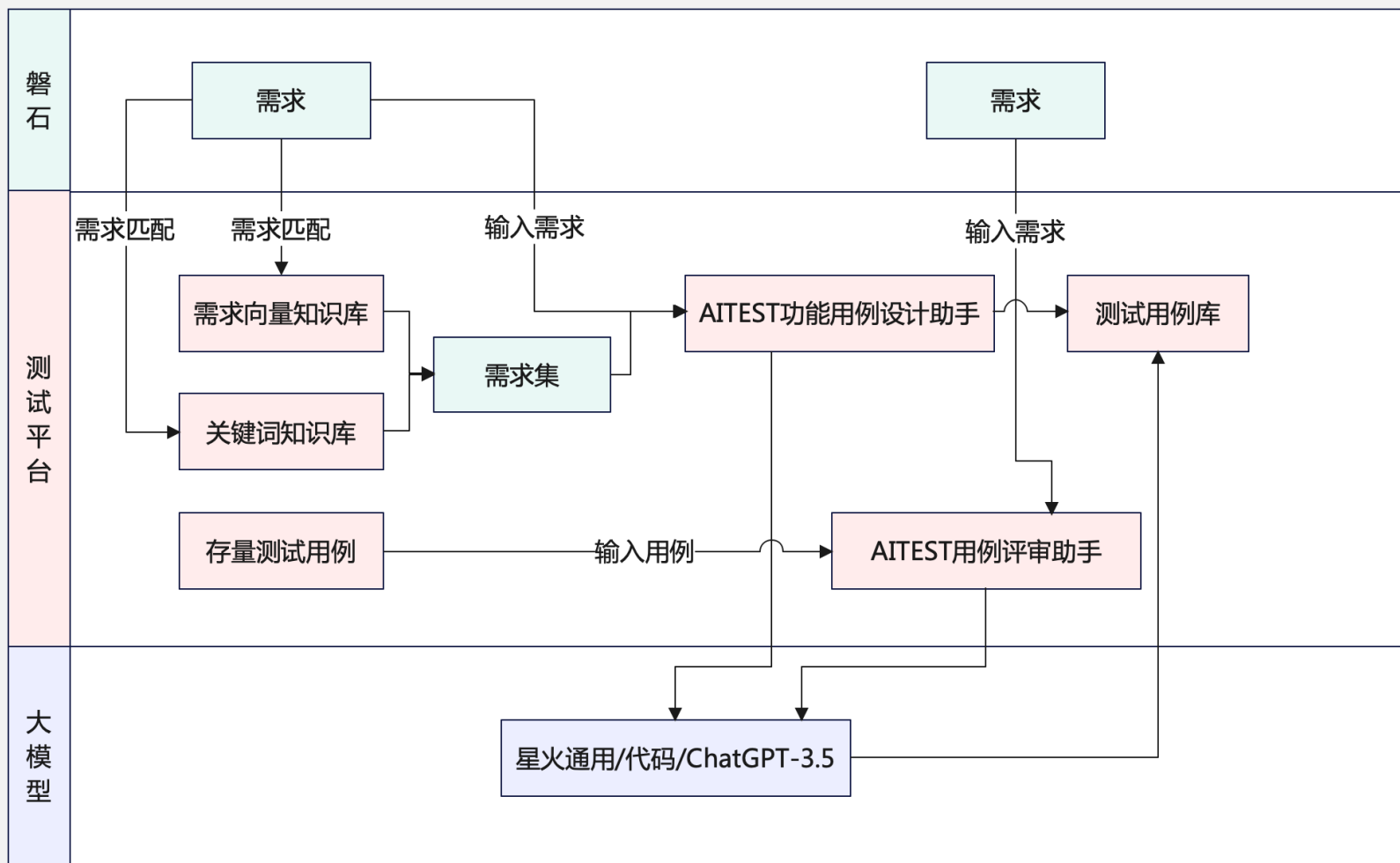


交互原理



未来展望——提升用例设计效果

- 建立需求的向量知识库及关键词知识库
- 完善测试平台功能用例设计助手，添加匹配知识库逻辑以完善需求，完善助手交互设计
- 增加需求评审助手，以存量用例及关联需求为输入，完善并补充测试用例



AI赋能接口测试的行业标杆——华为CodeArts-ATGen，可实现面向接口上下文的全场景测试智能生成。华为云ATGen基于正向API接口定义文档，自动挖掘API测试上下文操作依赖图（ODG），探索遍历ODG自主生成API测试序列，实时构造并下发API测试请求，判定API测试响应结果，动态修正ODG图，优化下一轮生成

创建任务

[illegible]

- ✓ 导入Yaml接口定义文件
- ✓ 配置环境和鉴权方式

- ✓ Yaml自动解析
- ✓ 导流数据自动解析

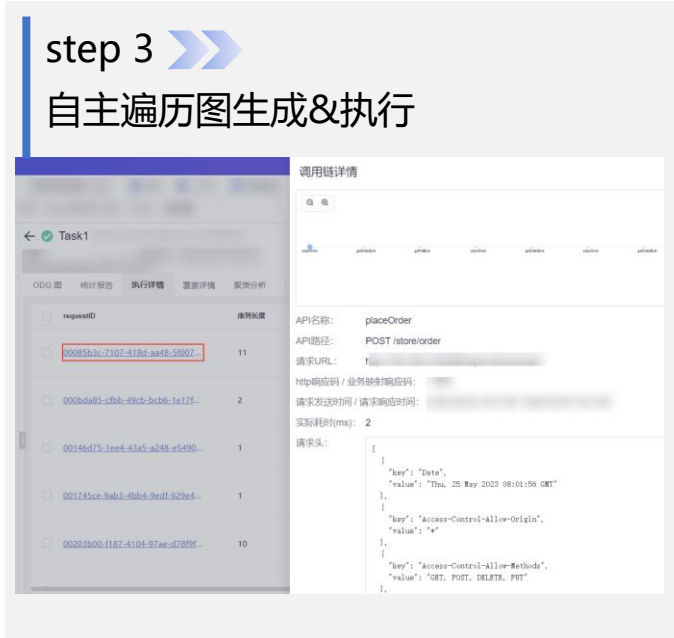
生成操作依赖图



- ✓ 可按需人工修改

- ✓ 自动生成操作上下文依赖ODG图

自主遍历图生成&执行

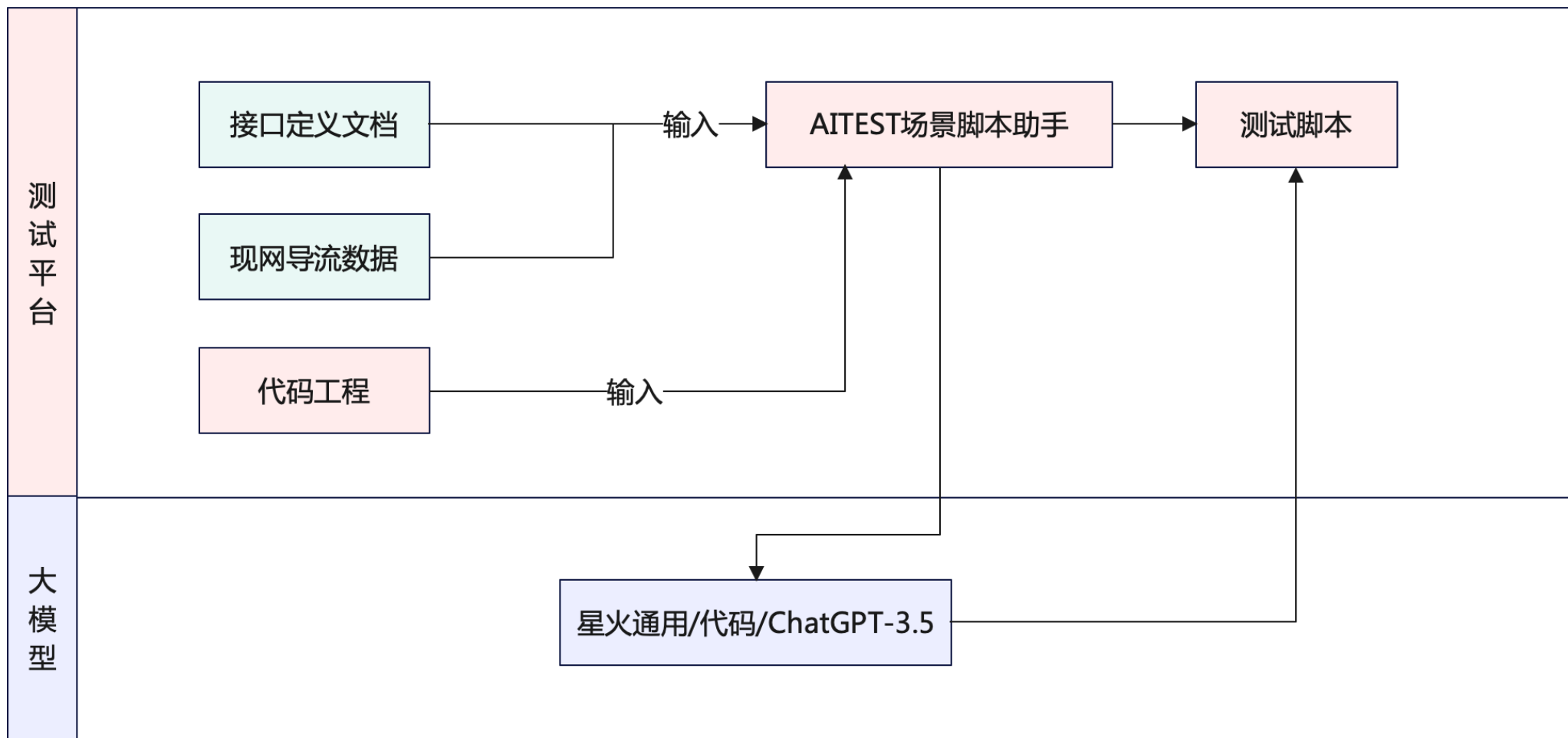


- ✓ 点击执行

- ✓ 自动化遍历图生成测试用例
- ✓ 自动生成测试数据
- ✓ 自动下发执行&收集反馈

未来展望——面向接口上下文的全场景测试智能生成

- 通过将现网导流数据结合接口定义文档输入给大模型，分析接口间的调用逻辑，提升场景脚本的准确性
- 通过代码大模型分析业务服务代码工程，分析接口间的调用逻辑，提升场景脚本准确性



交流讨论