

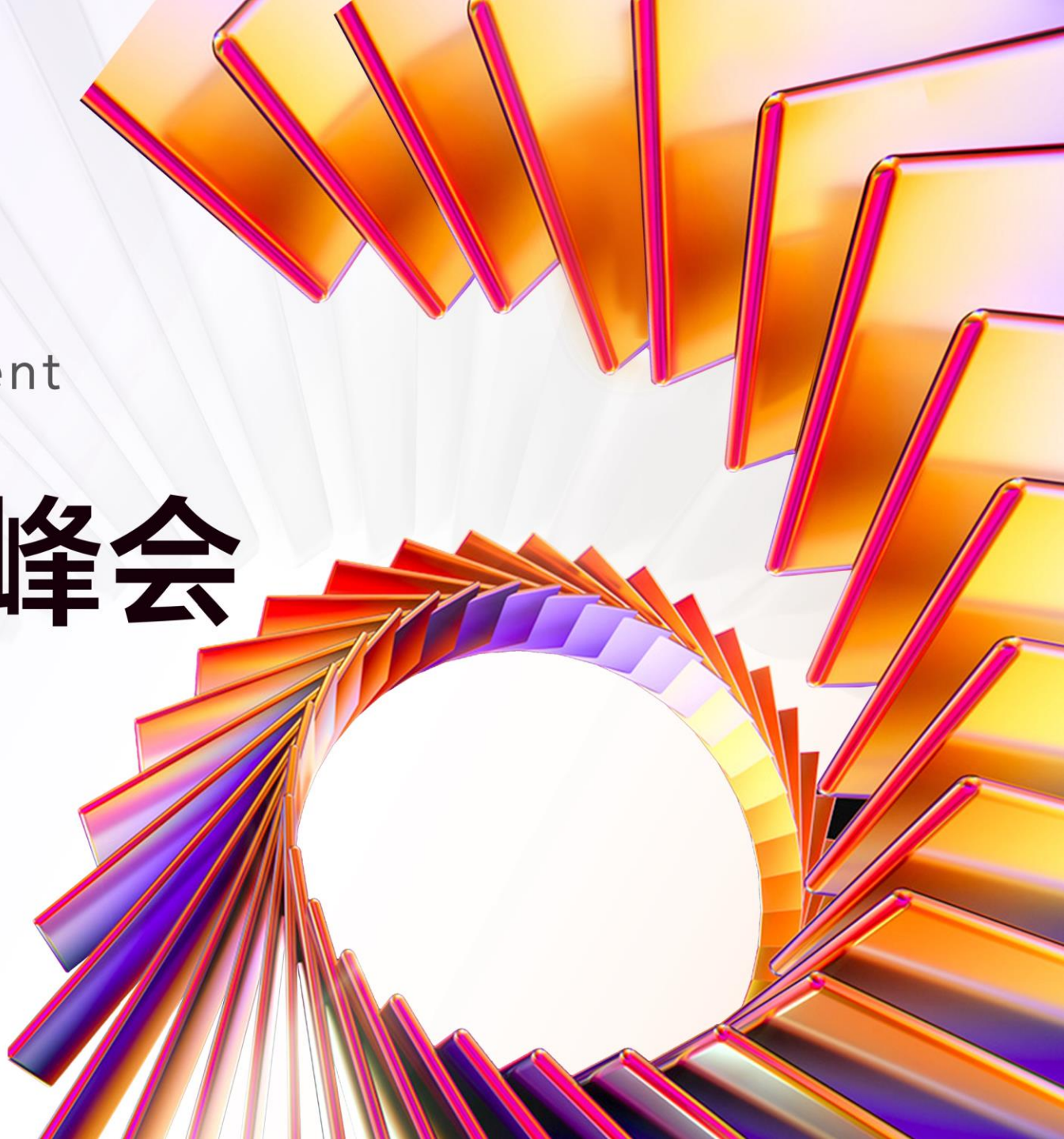


第7届 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发

8月8-9日 | 北京站





第8届AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

11/14-15 | 深圳站

12/19-20 | 上海站



查看会议详情

深圳站论坛设置

智能装备与机器人

超越“编程 Copilot”

下一代知识工程

智能网联与汽车智能化

AI 测试工具开发与应用

AI 基础设施和运维

数据智能及其行业应用

可信 AI 安全工程

大模型和 AI 应用评测

多 Agent 协同框架

从智能测试到自主测试

大模型推理优化

多模态 LLM 训练与应用

智能化 DevOps 流水线

上下文工程

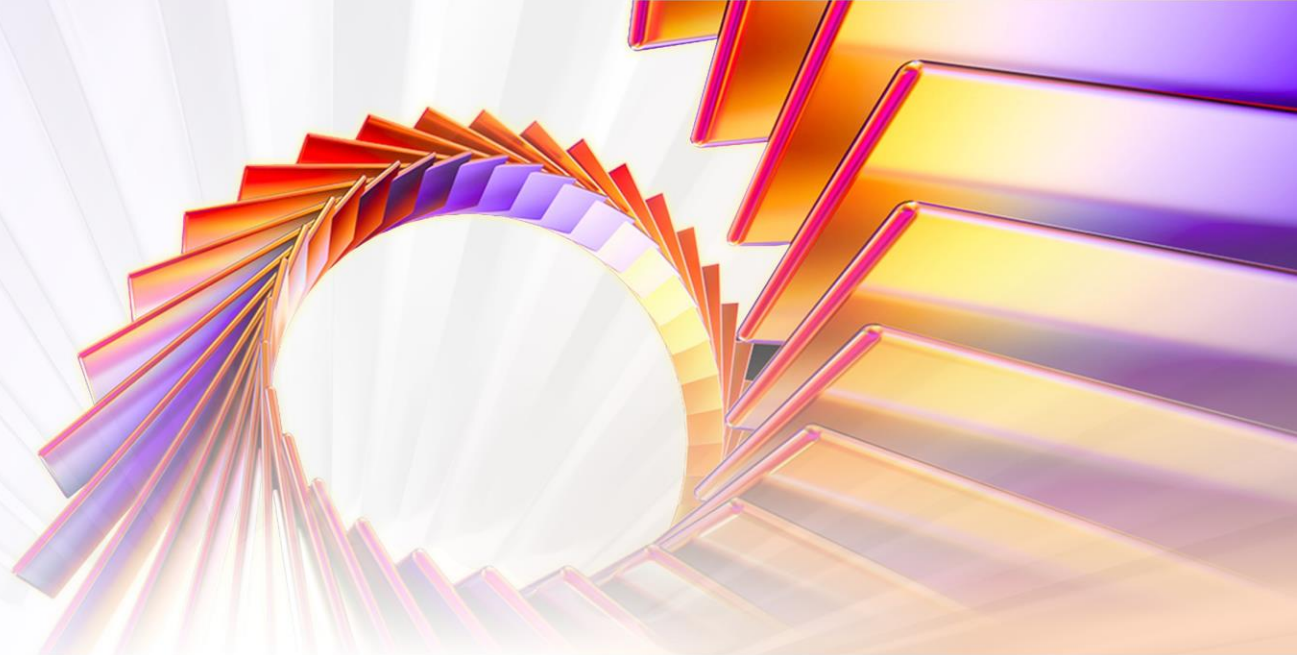


| 8月8-9日 | 北京站

第7届 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



基于大模型的文本测试用例生成 探索与实践

黄修添 | 字节跳动



黄修添

字节跳动-质量工程智能团队 算法专家

清华大学硕士毕业，曾就职于蚂蚁集团支付宝技术部及研发效能部。现任字节跳动算法专家，多年来深耕人工智能领域，专注于NLP与多模态等技术在质量与效率领域的融合创新。主导多个0到1的质量效能类项目的研发和落地，擅长利用 AI 驱动业务质量和效率的提升。

目录

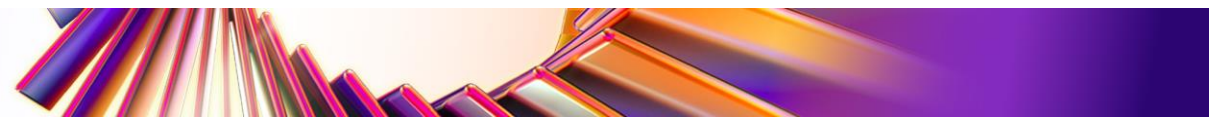
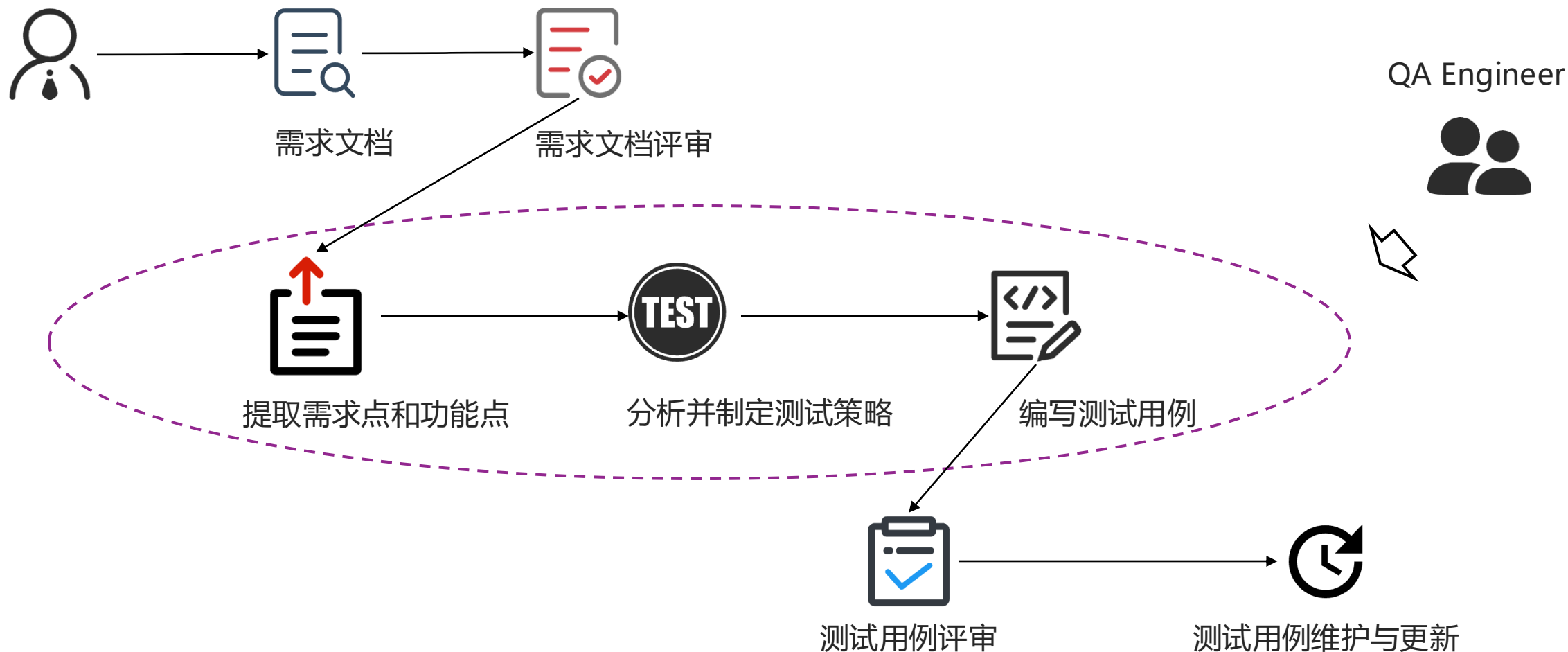
CONTENTS

- I. 背景与目标
- II. 技术方案
- III. 落地效果
- IV. 未来展望

PART 01

背景与目标

项目背景 需求驱动测试用例设计



业务需求日益繁杂

- 业务规模与复杂度快速提升
- 用例同步更新与维护压力大

软件质量保障难度大

- 用例依赖大量人工编写，耗时耗力
- 难以穷举边界条件，遗漏风险高
- 用例质量参差不齐，主观影响大

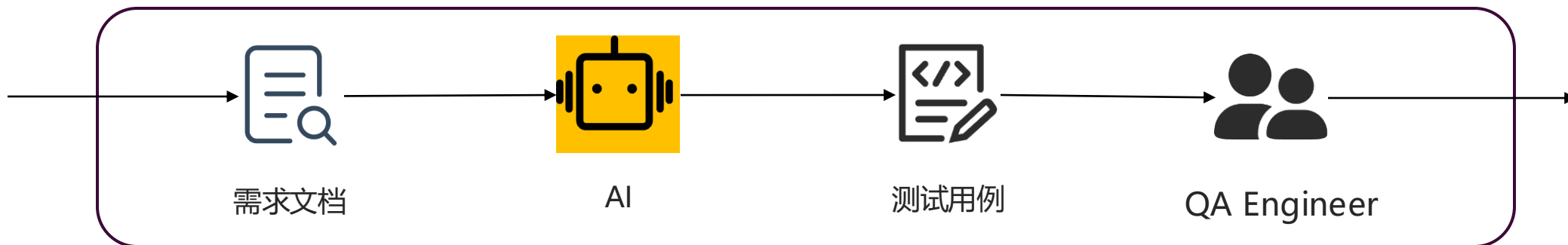
业务提质效诉求迫切

- 测试用例编写与维护成本高
- 测试用例覆盖率/质量不足
- 需求变更和上线响应不足

技术变革契机

- 大模型基础理解能力持续提升
- 多源多输入数据理解能力增强





降低用例编写成本

减少测试人员手动编写测试用例的时间和资源投入

提升用例生成质量

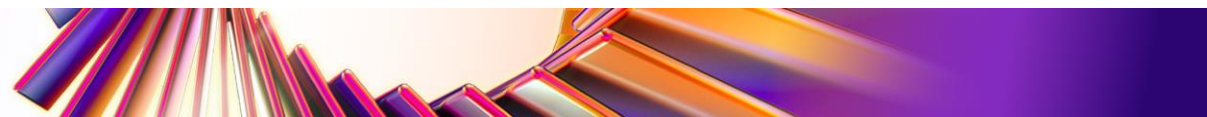
自动生成高质量、高覆盖率的测试用例，更贴合复杂业务实际

加速测试流程、提升研发效率

测试用例自动生成，实现测试提速，支撑敏捷持续交付

支撑多场景业务诉求

满足中、英文等多语言和多业务场景的测试用例需求





需求文档理解难度大

- 需求文档中包含图片和表格等多模态数据
- 需求上下文信息依赖跨度可能很大
- 需求文档质量参差不齐，业务差异大



训练数据采集困难，评测难度大

- 不同业务线存在风格、粒度差异，简单拼接数据效果有限
- 高质量、标准化的训练用例 / 需求对齐数据极其稀缺，采集难
- 评测标准不统一，人工评测难度大



用例准确率和覆盖度不足

- 需求文档有效信息覆盖有限，缺失业务经验知识
- 边界场景考虑不足，容易遗漏组合测试场景
- 业务用例风格差异大，生成用例难以满足



业务复杂诉求多

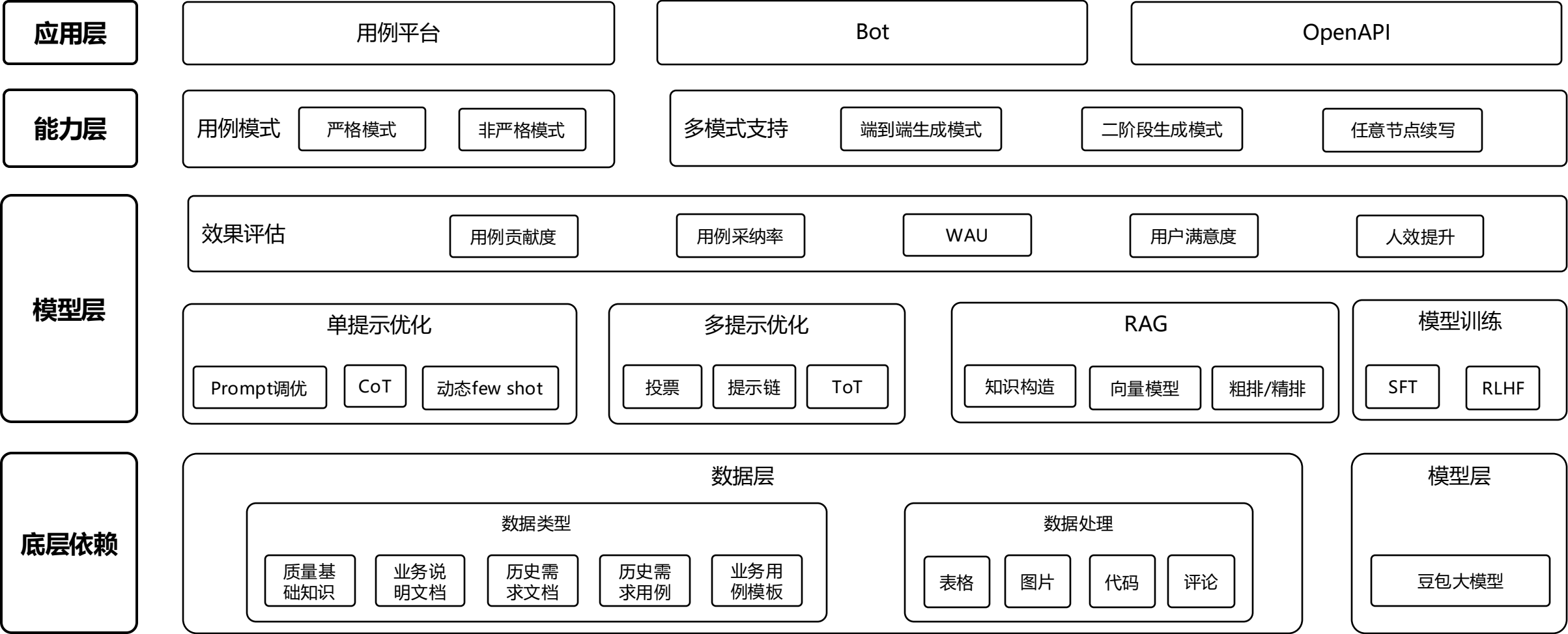
- 用例生成需兼顾多语言（中 / 英 / 其他）
- 多文档输入，如产品需求、技术需求、其他文档
- 支持业务指定用例模板生成测试用例
- 多种生成模式，覆盖端到端、续写模式等



PART 02

技术方案

整体建设思路



▶ 关键技术

需求文档解析

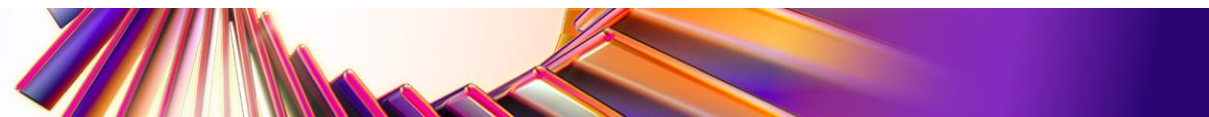
提示词工程

SFT & DPO

多文档 & 模板

RAG

Workflow

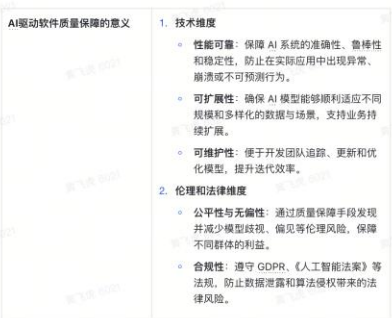


	这是标题1	这是标题2	这是标题3
对照组	配置信息a1	配置信息b1	配置信息c1
实验组1	配置信息a2	配置信息b2	配置信息c2
实验组2	配置信息a3	配置信息b3	配置信息c3

表头解析



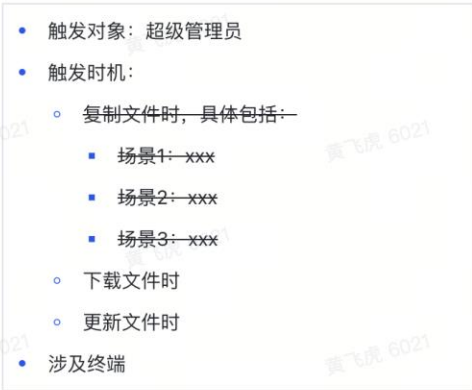
表格单元合并



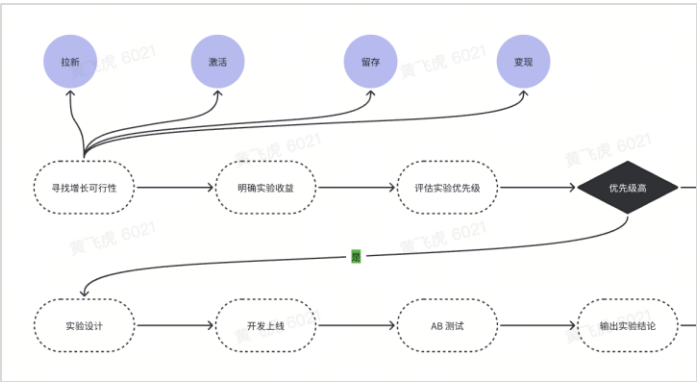
表格内容层级解析



图片链接/布局解析



无效内容



流程图解析



需求背景

现状

- 复制文件：将文件从A复制到B
- 下载文件：将文件从A下载到B
- 更新文件：将文件从A直接更新



复制文件



下载文件



更新文件

目前，XX软件的用户账号登录流程存在如下问题：

- 登录步骤多，操作繁琐，首次登录需输入账号、密码和验证码，部分场景下还需二次验证；
- 支持手机号、邮箱登录，但没有统一入口，用户易于混淆；
- 登录失败后的错误提示不够清晰，缺乏针对性操作建议；
- 登录成功后未做多端登录提醒，部分用户存在信息安全隐患；



需求背景

现状

- 复制文件：将文件从A复制到B
- 下载文件：将文件从A下载到B
- 更新文件：将文件从A直接更新

<!!image:DsFbbAmwlooivywxwyRlczZsOnbg!>

<!!image:LFTBbgLujo0ykbxcUrKcDGOynqd!>

<!!image:Eli9b7NUvoFCraxTBpDcVYEbn4d!>

目前，XX软件的用户账号登录流程存在如下问题：

- 登录步骤多，操作繁琐，首次登录需输入账号、密码和验证码，部分场景下还需二次验证；
- 支持手机号、邮箱登录，但没有统一入口，用户易于混淆；
- 登录失败后的错误提示不够清晰，缺乏针对性操作建议；
- 登录成功后未做多端登录提醒，部分用户存在信息安全隐患；

产品方案

功能可用版本

版本	该版本是否可用此功能	该版本是否要按需另购此功能
基础版	☒	☐
专业版	☒	☐
企业版	☒	☐

功能上线地区

上线地区	
中国大陆地区	☐
港澳台及海外	☐



产品方案

功能可用版本

| 版本 | **该版本是否可用此功能** | **该版本是否要按需另购此功能** |

| -- | ---- | ---- |

| 基础版 | [x] | ☐ |

| 专业版 | [x] | ☐ |

| 企业版 | [x] | ☐ |

功能上线地区

| 上线地区 | |

| ---- | -- |

| 中国大陆地区 | ☐ |

| 港澳台及海外 | ☐ |



▶ 提示词工程 (PE) 常见的PE技术

单提示技术



xxx-shot



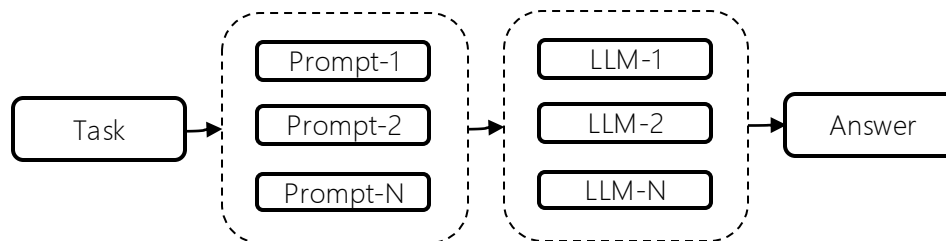
Chain-of-Thought



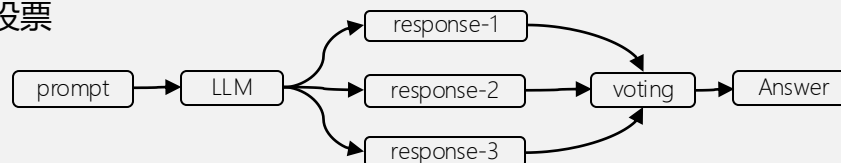
Program-Aided LM

理解问题并生成程序作为推理步骤

多提示技术



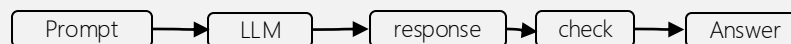
投票



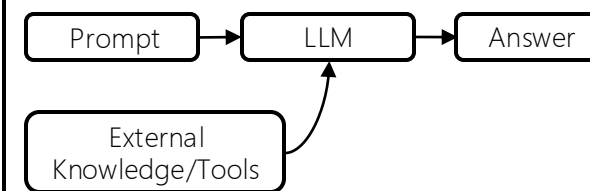
提示链



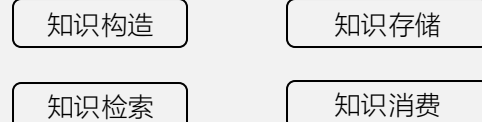
反思&检查 (如思维树)



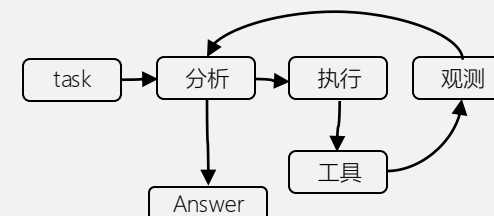
外部知识/工具



RAG

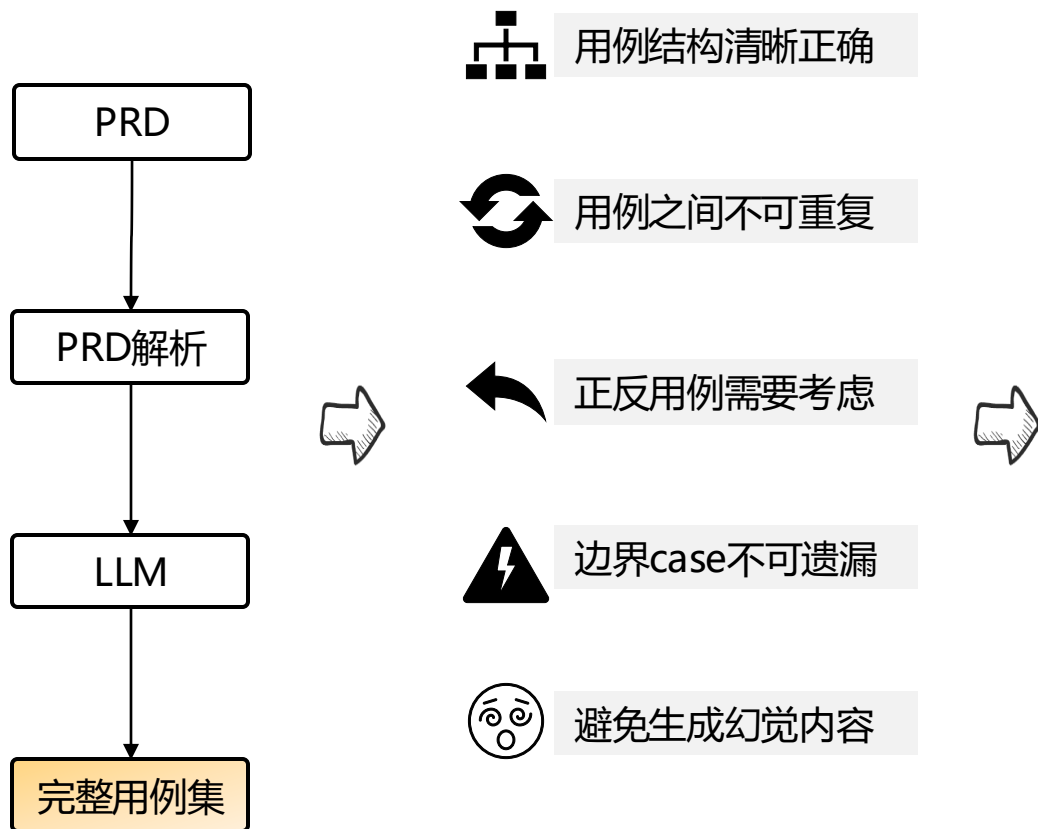


ReAct



▶ 提示词工程 (PE) 关键实践

端到端：输入需求文档，生成完整用例集



【角色提示】

<Role>
你是一个资深测试专家。
</Role>

- 指定风格：调整输出风格和语气
- 提升专注度：保障模型在特定场景的效果
- 提高准确率：充分发挥模型已具备的知识（如法律）

【目标清晰】

<object>
已知用户输入产品需求文档，请根据该文档生成文本用例集合。
</object>

- 明确目标，减少输出无效内容

【明确 workflow】

<workflow>
1. 阅读并理解产品需求文档，确保对产品功能和需求有全面了解。
2. 识别文档中涉及的所有功能模块，并生成相关测试功能点。
3. 依次分析每个测试功能点，生成相关测试用例。
</workflow>

- 明确工作步骤，减少模型理解成本
- 引入思维链，保障输出结果可靠性

【规范约束】

<constraint>
1. 每个测试功能点是独立的，和其他功能点不可重复。
2. 测试内容需要同时考虑正反用例，不可缺失和冗余。
3.
</constraint>

- 告诉模型不要做什么，减少非预期的输出

【规范输出格式】

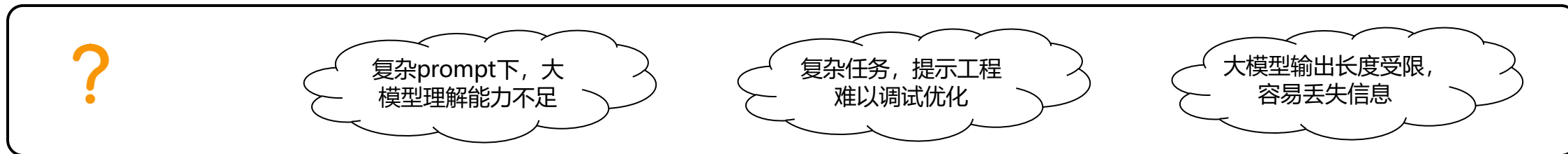
<output format>
平铺/层级结构
</output format>

【示例】

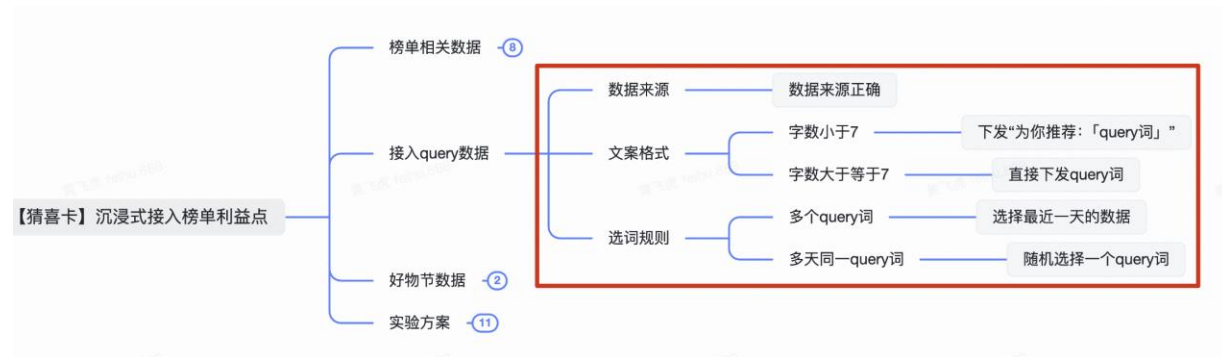
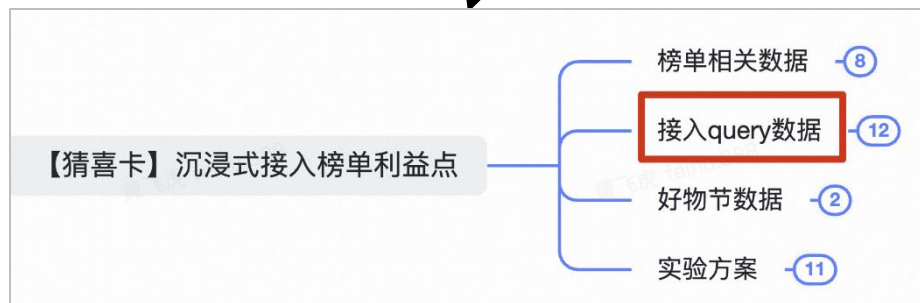
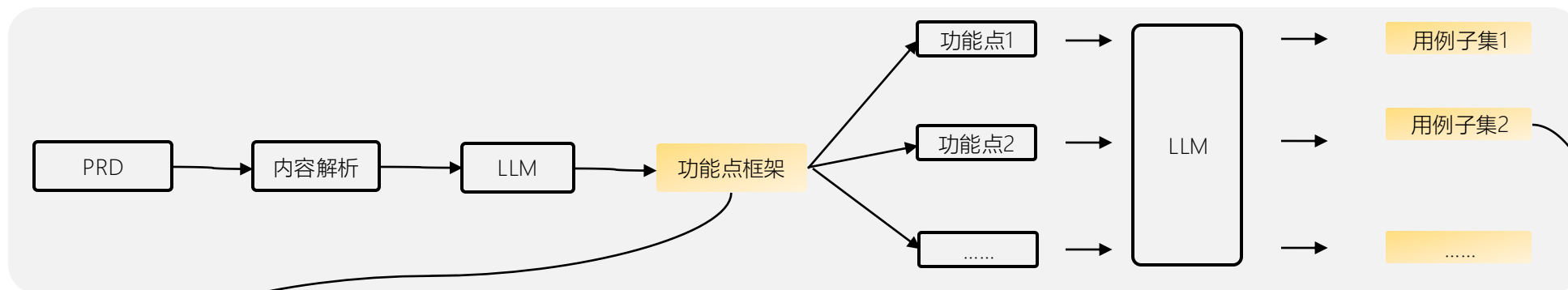
<examples>
.....
</examples>

- 准确性：减少对指令的误解
- 一致性：强制执行统一的结构和风格
- 性能：高质量示例提升处理复杂任务的能力
- 示例选择：相关的、多样的、清晰的

▶ 提示词工程 (PE) 关键实践



二阶段生成



① PE技术的不足

- 受限于预训练模型固有能力：模型未深度理解业务语境，效果依赖 Prompt 适配
- 易出现“指令幻觉”与一致性问题：相同需求，生成的用例细节差异大
- 复杂业务场景下效果存在天花板：难以引入业务特有规则、历史用例等深层知识

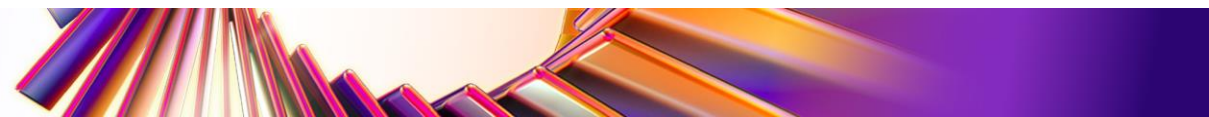


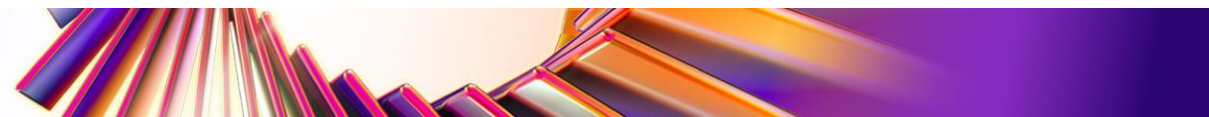
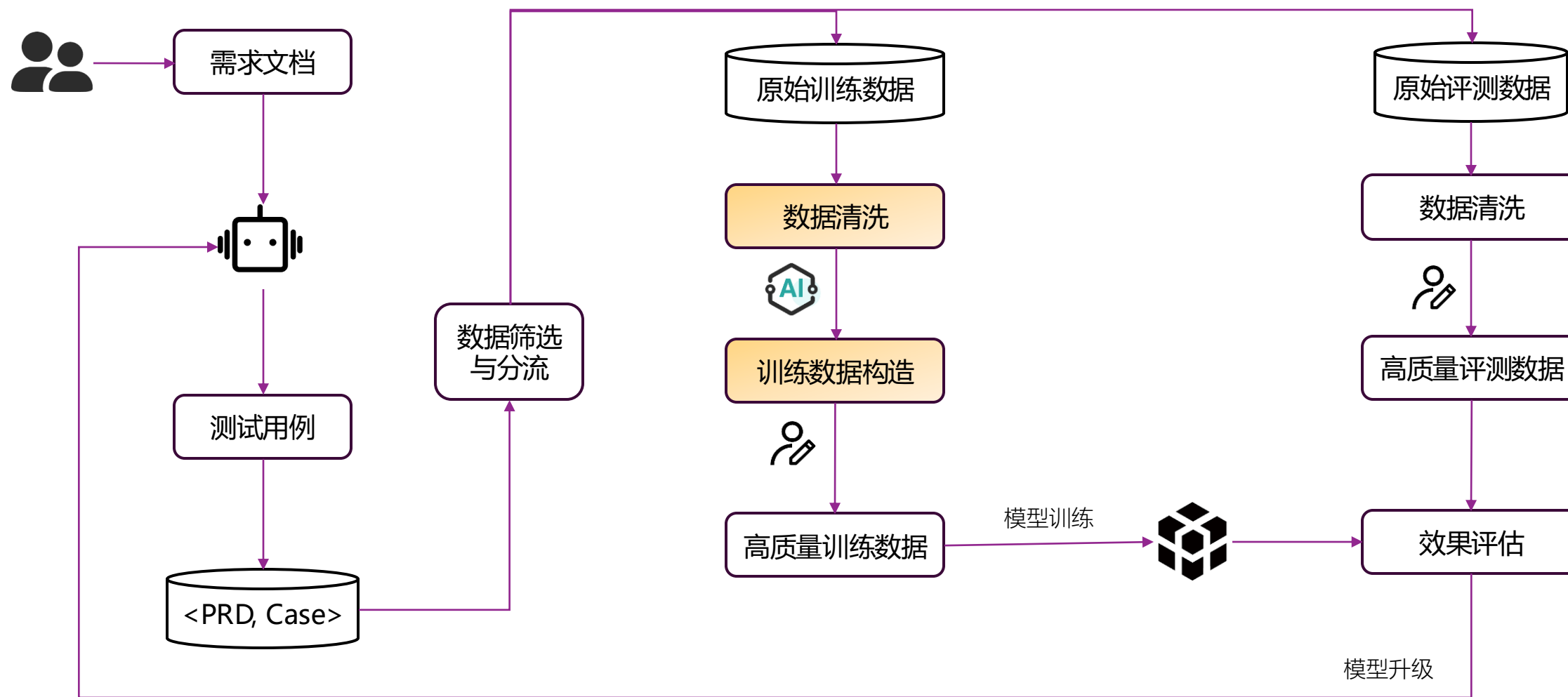
U 策略

- Prompt 为先，探索空间：快速收敛需求、发现瓶颈
- SFT 深耕，提升上限：在数据和 Prompt 之上，提升生成质量和泛化能力
- 两者结合，效果最大化：Prompt 持续探索 + SFT 周期性升级，形成闭环优化

🧠 SFT & DPO的优势

- 深度业务融合：整合历史用例、规范、领域知识，模型内化业务专有 能力
- 生成质量稳定：微调后可显著提升一致性、覆盖率与采纳率
- 支持复杂场景与多语言：针对具体需求和语言场景做定制化优化





需求背景

现状

- 复制文件：将文件从A复制到B
- 下载文件：将文件从A下载到B
- 更新文件：将文件从A直接更新



复制文件



下载文件



更新文件

目前，XX软件的用户账号登录流程存在如下问题：

- 登录步骤多，操作繁琐，首次登录需输入账号、密码和验证码，部分场景下还需二次验证；
- 支持手机号、邮箱登录，但没有统一入口，用户易于混淆；
- 登录失败后的错误提示不够清晰，缺乏针对性操作建议；
- 登录成功后未做多端登录提醒，部分用户存在信息安全隐患；
- 用户忘记密码重置流程不够便捷，易导致用户流失。

问题

- 新用户初次登录转化率低，流程不顺畅影响体验；



```
# 这是一篇测试需求文档
## 登录流程优化
### 统一登录入口
#### 登录方式入口合并
#### 输入框自动识别
#### 简化登录流程
#### 一次验证通过
##### 验证码方式选择
##### 登录流程完整性
#### 新用户注册后登录
#### 处于未注册账号状态
#### 忘记密码流程
### 登录安全性
#### 异常登录校验
#### 异地登录校验
#### 异常设备登录校验
```

SFT数据样例

```
# 这是一篇测试需求文档
## 登录流程优化
### 统一登录入口
#### 登录方式入口合并
#### 输入框自动识别
#### 简化登录流程
#### 一次验证通过
```

DPO数据样例

```
# 这是一篇测试需求文档
## 登录流程优化
### 统一登录入口
#### 登录方式入口合并
#### 输入框自动识别
#### 简化登录流程
#### 一次验证通过
##### 验证码方式选择
##### 登录流程完整性
#### 新用户注册后登录
#### 处于未注册账号状态
#### 忘记密码流程
### 登录安全性
#### 异常登录校验
#### 异地登录校验
#### 异常设备登录校验
```



多文档&模板

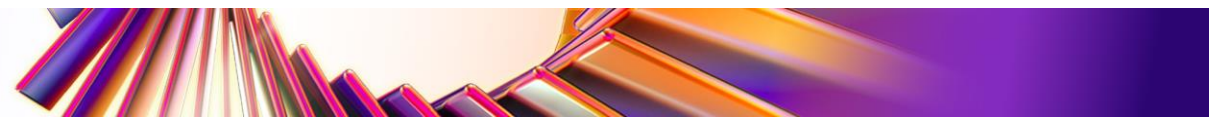
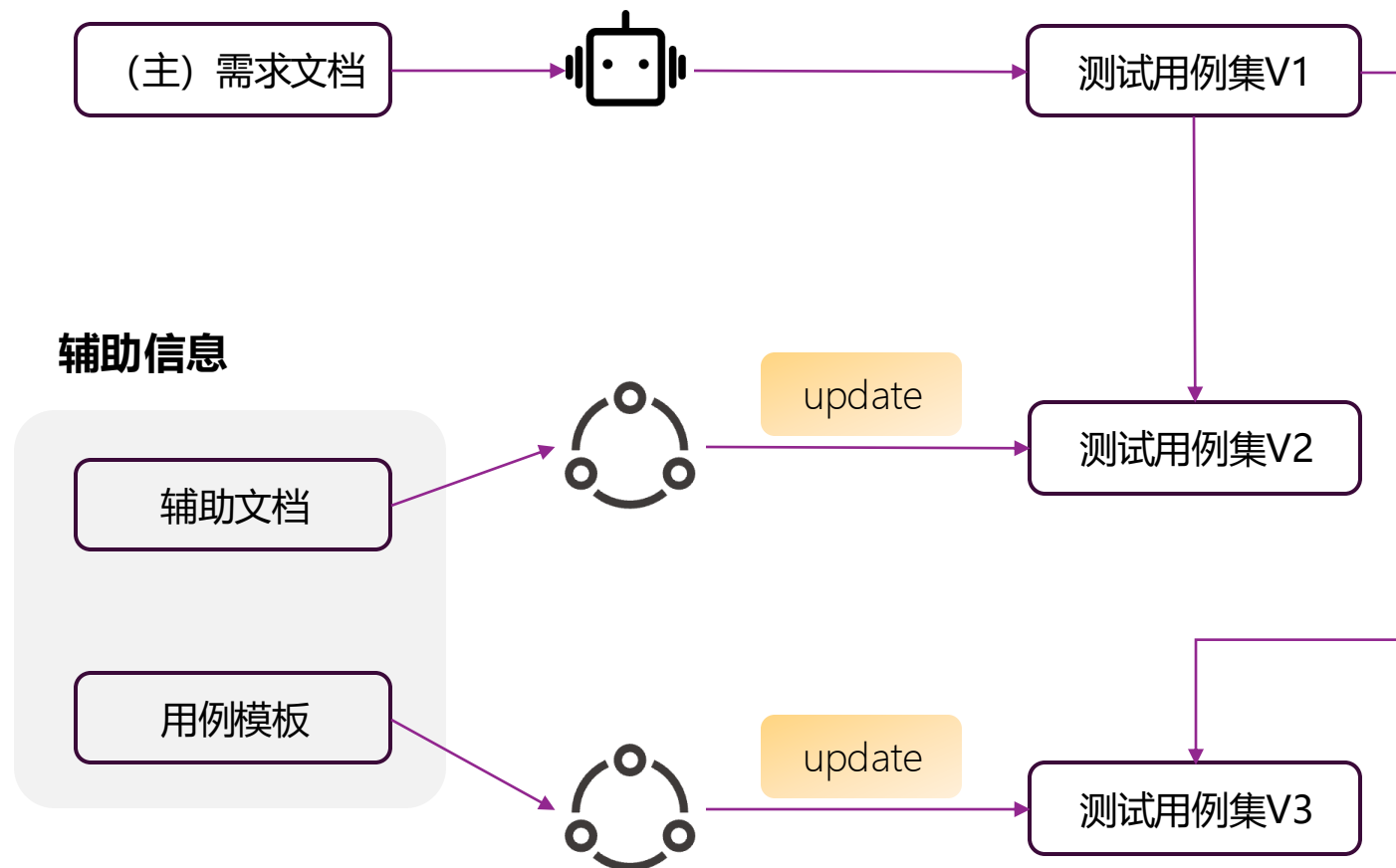
业务复杂场景

- 系统涉及多个业务线、模块，需求分散于多份文档（产品需求文档、技术需求文档、埋点文档等）
- 传统单文档编写存在信息片面、覆盖盲区等问题

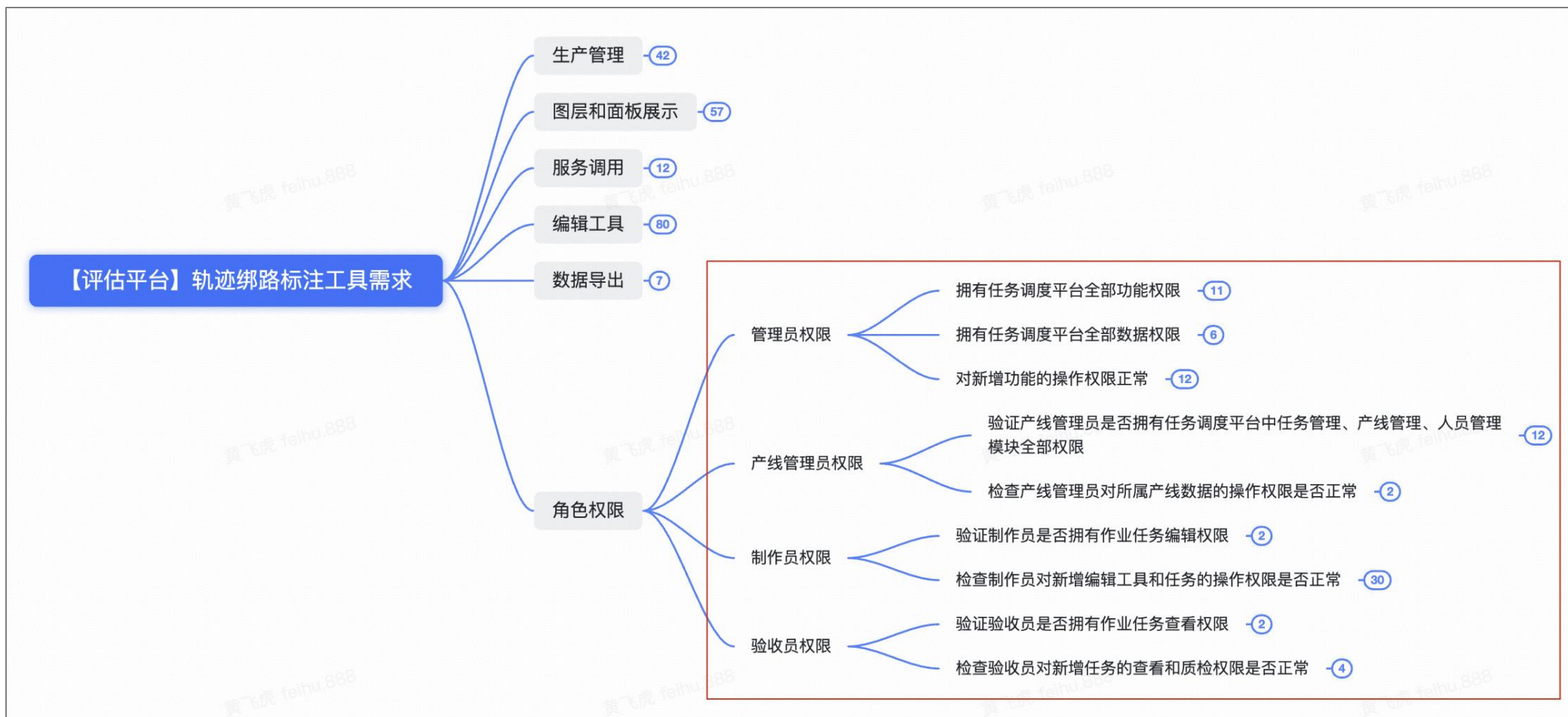


高“贴合度” + “定制化”需求

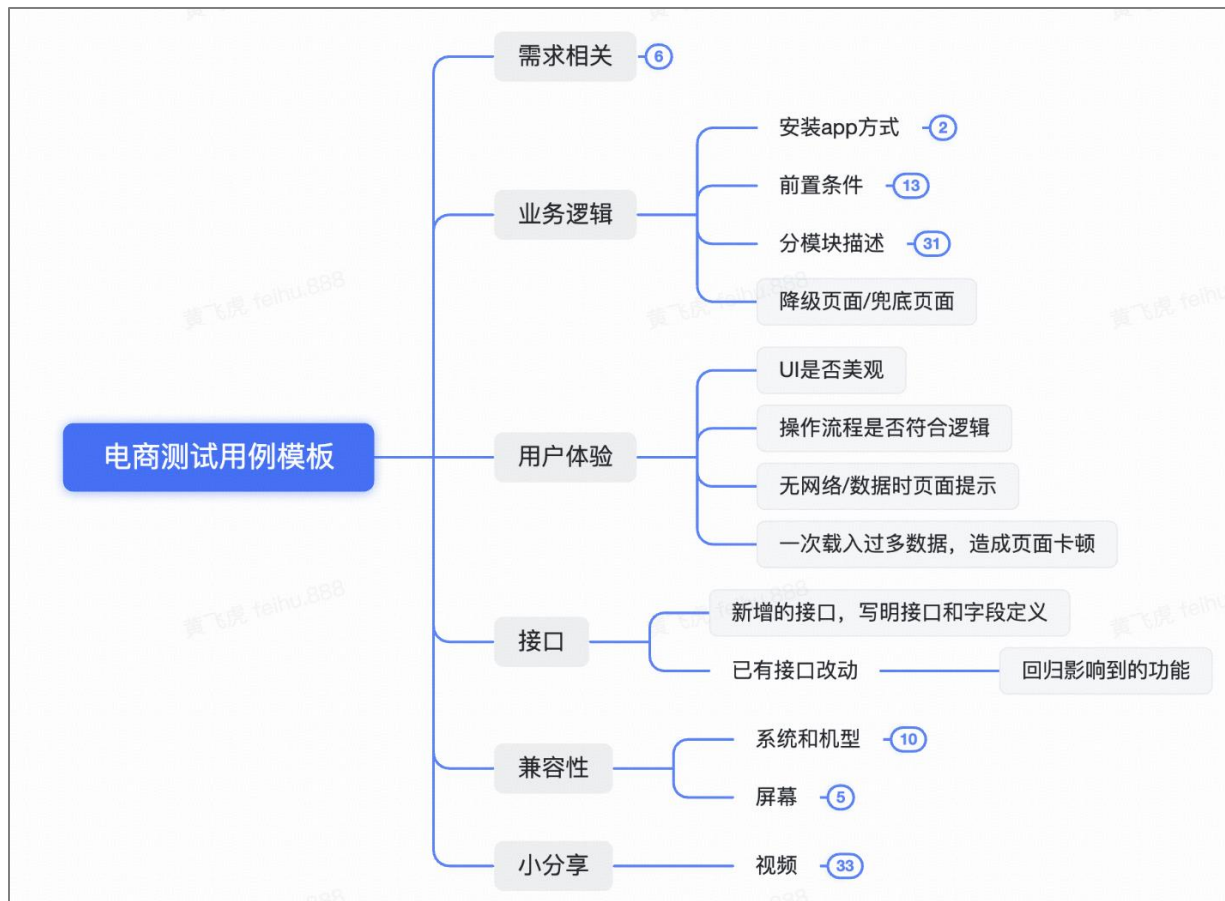
- 各业务有特定测试要素、场景偏好、输出结构
- 需支持不同行业、团队定制自己的用例模板规范



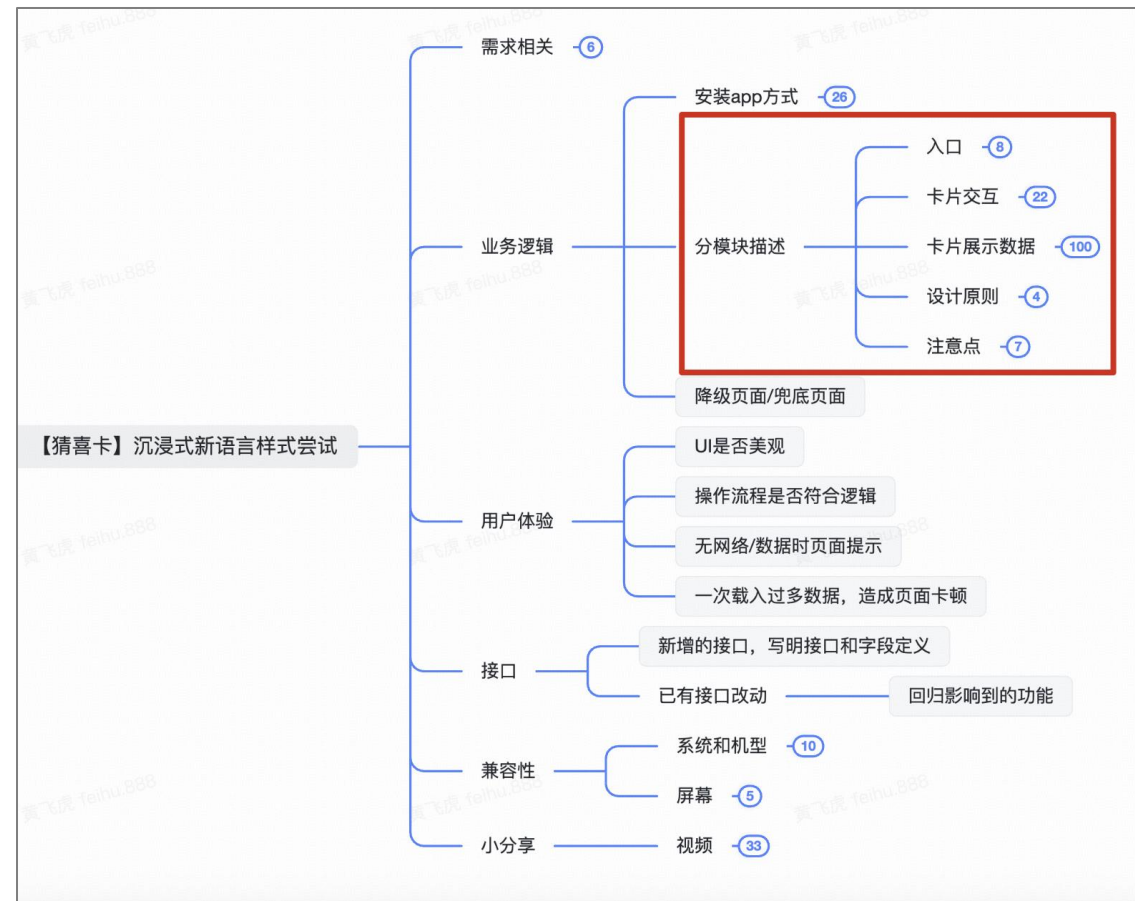
需求文档：《轨迹绑路标注工具需求》+ 《角色权限说明》



多文档&模板



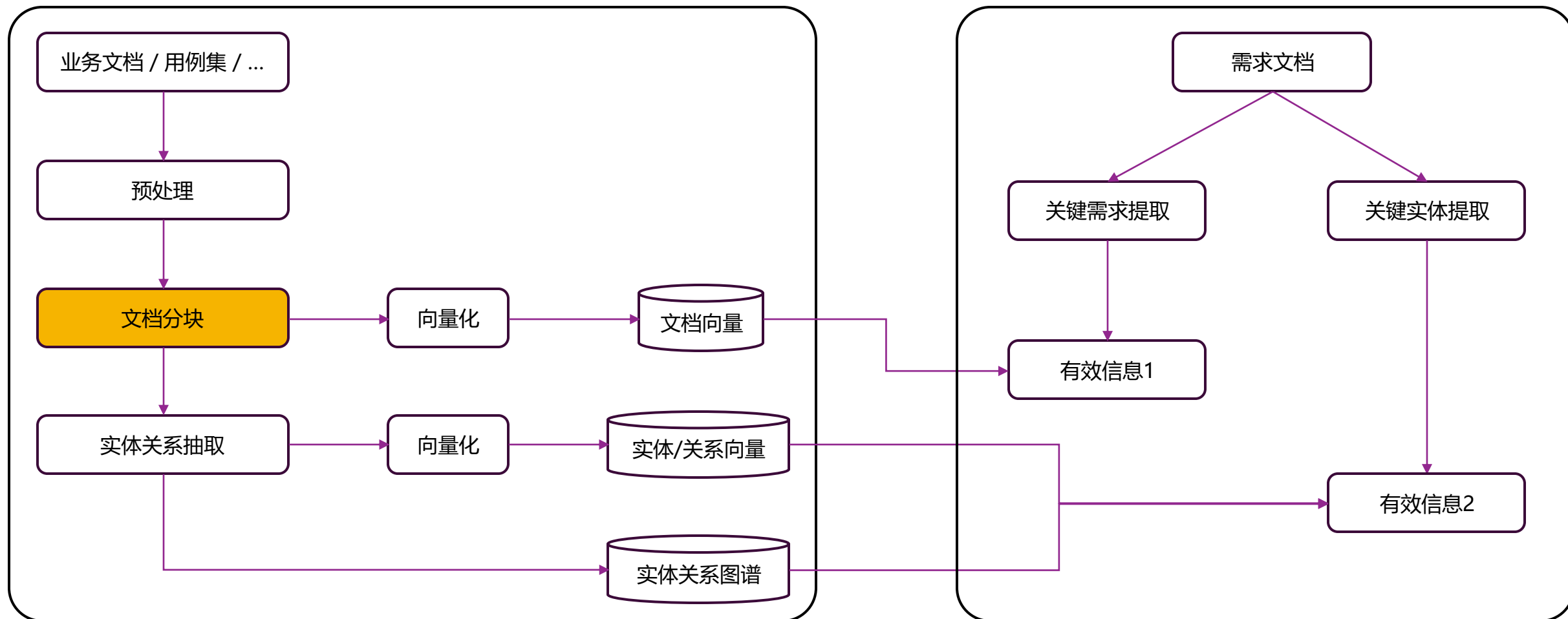
测试用例模板

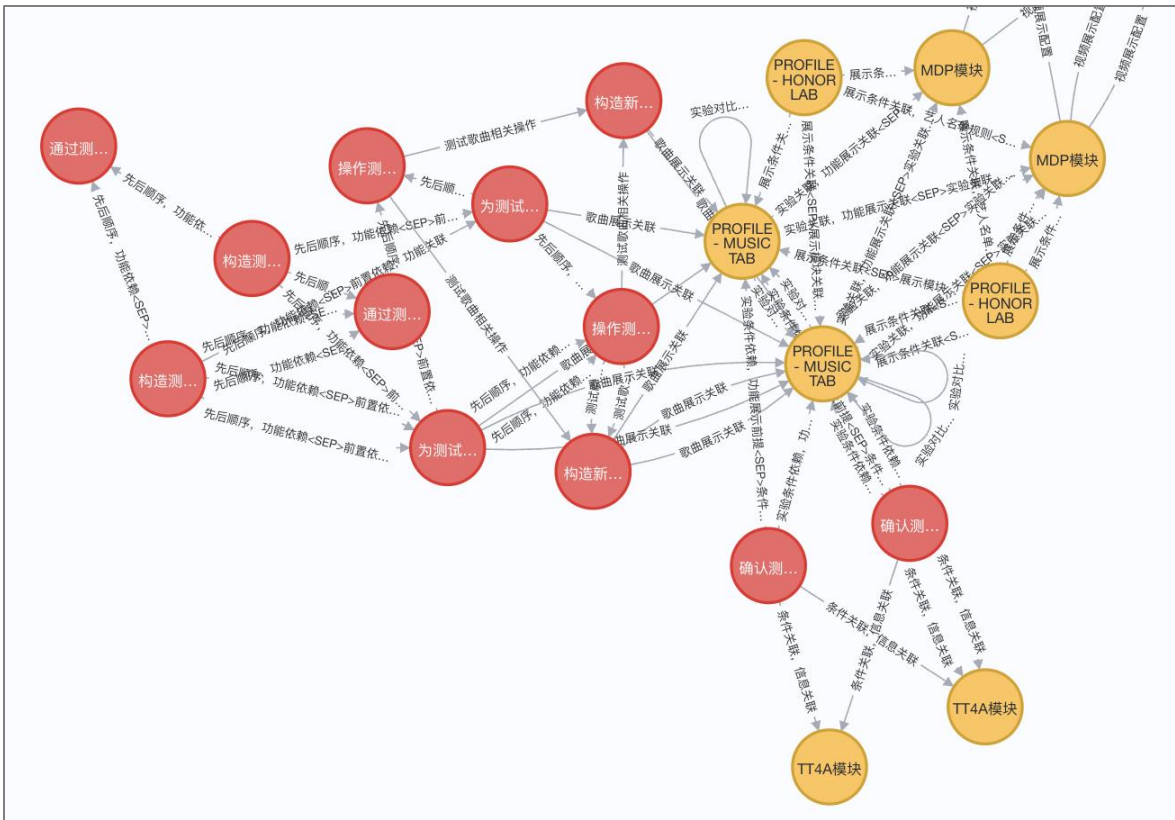


《【猜喜卡】沉浸式新语言样式尝试》用例



核心思想：Vector-based RAG + Graph-based RAG





模块/需求关联知识图谱示例

< ecommerce_duanshipin

原始文档 知识问答 Config Prompt 知识检索

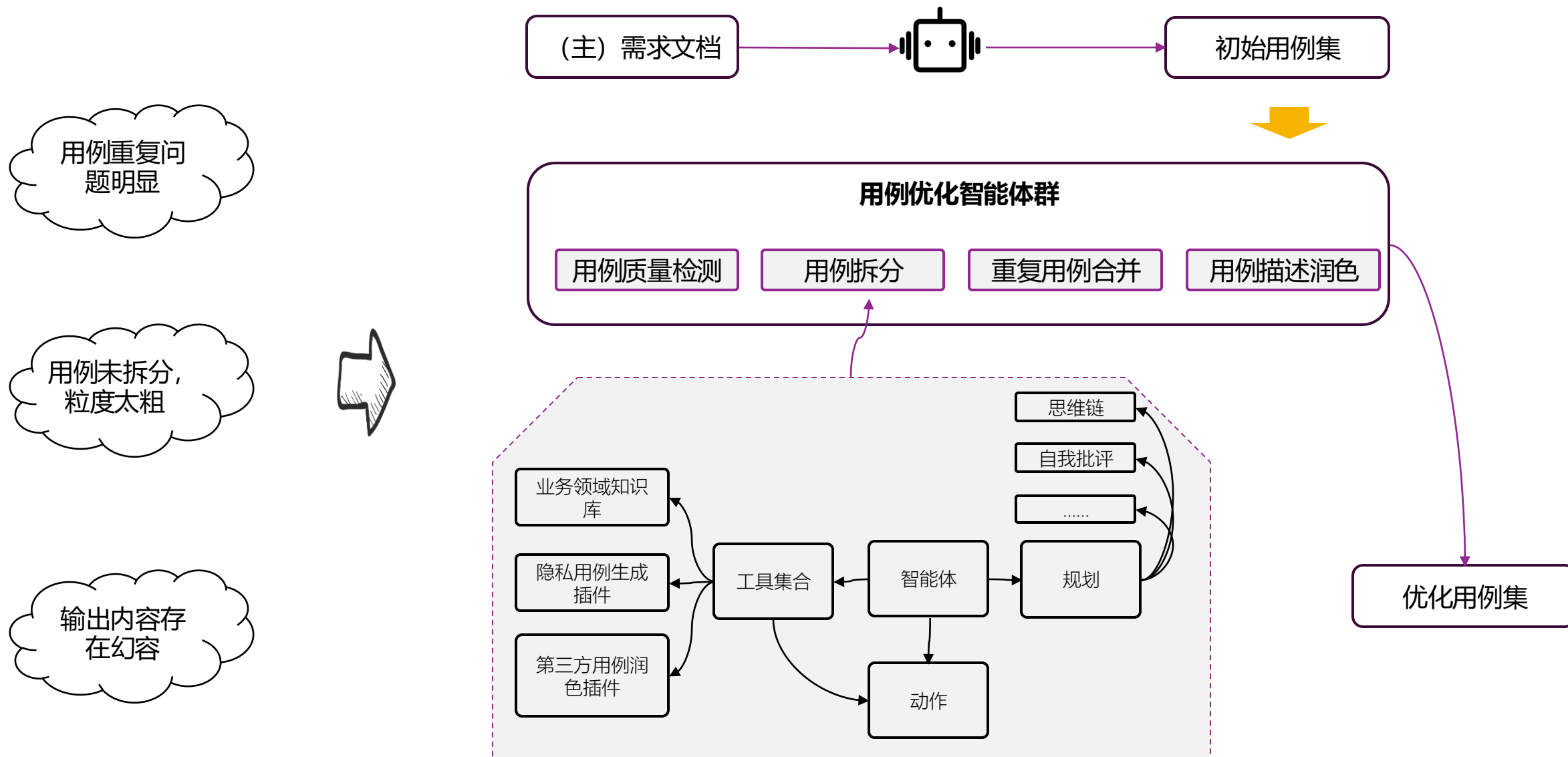
+ 导入文档

文档名称	文档ID	文档状态
文档16	03da1d50c69de58a6a0cefeccb3d5f300b8289c1693b3d42680f308c3da53c57	PROCESSED
文档32	17791cfc161b64b05262cfbdf188d73ab19d0a42e7f17a8d30b80e9b30c60f2a	PROCESSED
文档24	1987e9fe8ff9c1a90192431f2158e9141ca4f920de40c1711eb42a29c2dd69cb	PROCESSED
文档11	1b150aad255974c0fd22cf56ea8519253bfe3d88ebd52036fc7f578e26b1fd76	PROCESSED
文档9	1b2d49e331285bb83f2f1bf9cea616f4beeb10cc9a0551d1827e35d2d89cbf2a	PROCESSED
文档17	1e20219fe31edfd9207d211fc71742083afc5845187f922a7a2d5bf30d0c0e13	PROCESSED
文档3	20a3b2387c7fbf1cd4c82e5de906b6270eb99afeeb115860c4a2a11dd7b108de	PROCESSED
文档13	24c560af4408af636f8350e8dae1d527f7f51e368bac5557610e512115561de1	PROCESSED
文档28	260c1c635bc24b25096655921837b86dd478249c32ff2dae1d4f95a04694bbd6	PROCESSED
文档20	26daadd977cbacc8d33ac52d70407bc9a36a946d9cccd414502984ad6d46f1b8	PROCESSED

向量知识库示例



Workflow

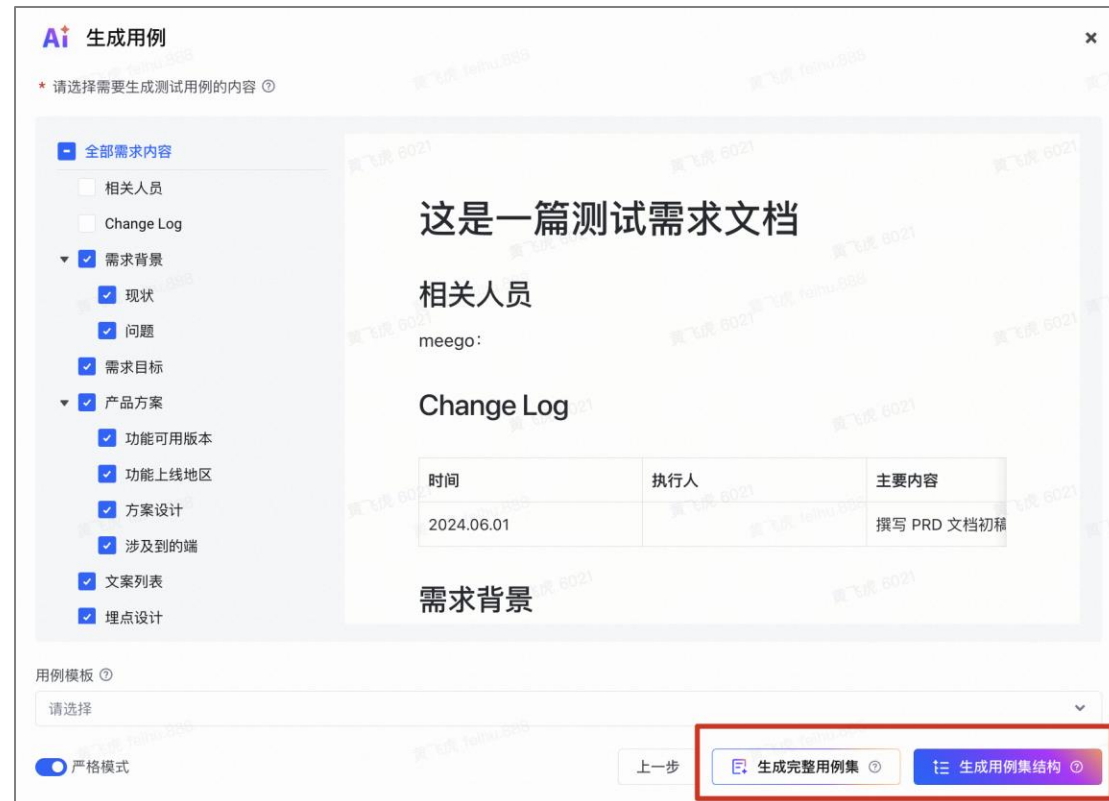


PART 03

落地效果

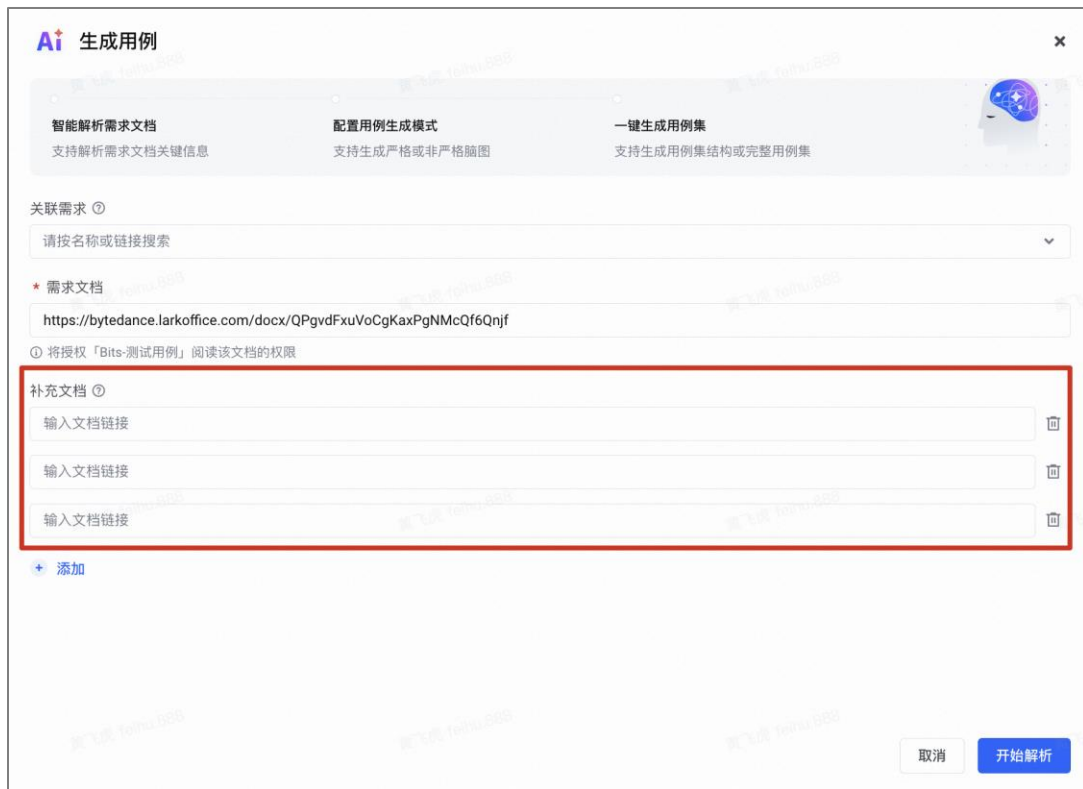


支持内容筛选



多种模式选择





多文档能力



模板能力





智能用例续写



用例分析与标注

落地效果 (~25Q1)

2.4W+

累计使用需求数

25%+

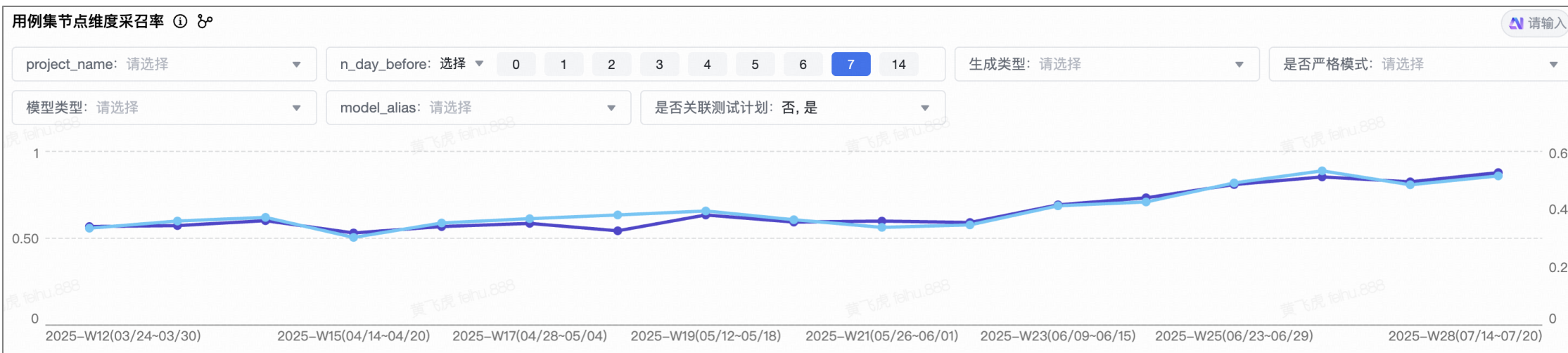
实际业务投产提效

18%+

AI用例集渗透率

50%+

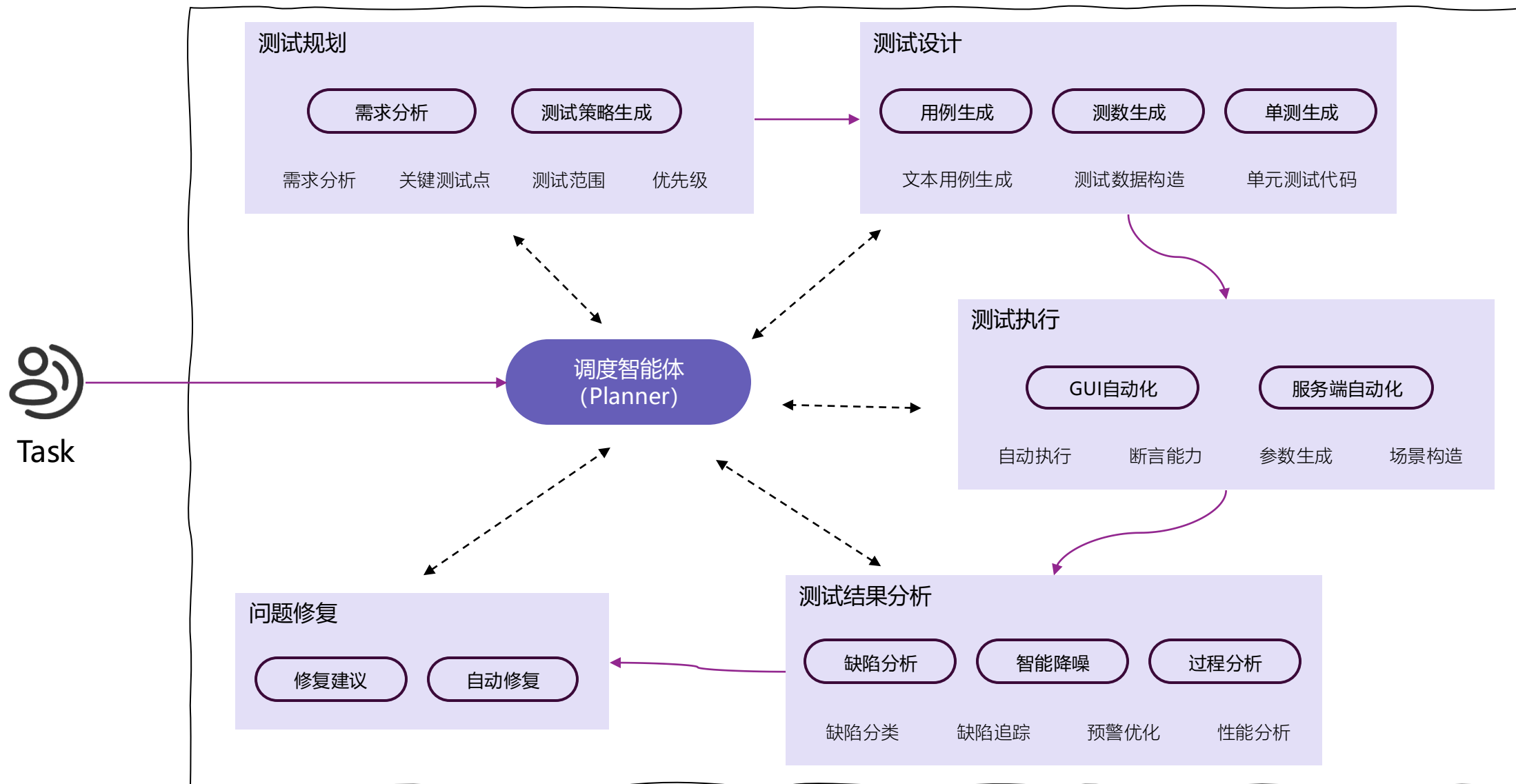
用例节点采纳率/召回率



PART 04

未来展望

未来展望





第8届AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

11/14-15 | 深圳站

12/19-20 | 上海站



查看会议详情

深圳站论坛设置

智能装备与机器人

超越“编程 Copilot”

下一代知识工程

智能网联与汽车智能化

AI 测试工具开发与应用

AI 基础设施和运维

数据智能及其行业应用

可信 AI 安全工程

大模型和 AI 应用评测

多 Agent 协同框架

从智能测试到自主测试

大模型推理优化

多模态 LLM 训练与应用

智能化 DevOps 流水线

上下文工程

AiDD

「深行 · 浅智」

Walk Deep, Think Light.

2025.11.16

AiDD首届麦理浩径徒步





科技生态圈峰会 + 深度研习

——1000+ 技术团队的选择



AiDD峰会详情





第7届AI+研发数字峰会
AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

