



AI+ 研发数字峰会
AI+ Development Digital summit



AI驱动抖音用户体验中台 探索与实践

杨昕 | 抖音集团-智能服务

科技生态圈峰会 + 深度研习



—1000+ 技术团队的选择



 **K+峰会**  **敦煌站**

K+ 思考周®研习社

时间: 2025.08.29-30

 **K+峰会**  **上海站**

K+ 金融专场

时间: 2025.10.17-18

 **K+峰会**  **香港站**

K+ 思考周®研习社

时间: 2025.11.25-26



K+峰会详情



 **AiDD峰会**  **上海站**

AI+研发数字峰会

时间: 2025.05.17-18

 **AiDD峰会**  **北京站**

AI+研发数字峰会

时间: 2025.08.08-09

 **AiDD峰会**  **深圳站**

AI+研发数字峰会

时间: 2025.11.28-29



AiDD峰会详情



杨昕

抖音用户体验中台后台技术负责人

在字节跳动负责抖音用户体验中台建设，曾在腾讯用户研究与体验设计部门负责建设反馈中台，调研云平台的建设。

目录

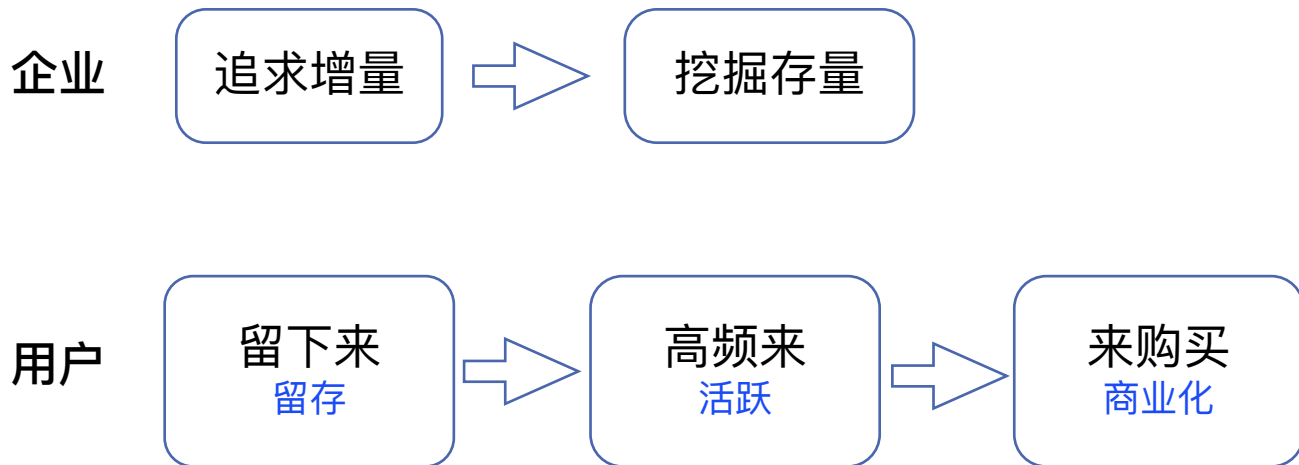
CONTENTS

1. 抖音用户体验中台的背景
2. 阶段一：引入AI工具来看清楚体验问题
3. 阶段二：大模型的工具化到平台化
4. 总结与展望

PART 01

抖音用户体验中台的建设背景

► 为什么要建设用户体验中台



从互联网红利时代转变到体验经济时代

背景介绍

- 1、在互联网大盘增长趋缓的环境下，以及抖音在用户大盘的渗透逐渐趋于饱和，特别是移动端用户渗透接近 **80%**，重点从追求增量转移到挖掘存量，提升用户活跃度上面。
- 2、在产品层面 [**覆盖全人群、服务场景更丰富、更不可或缺、社会价值更高**] 是抖音未来3~5年重要使命和目标。
- 3、如何及时发现不同用户的场景化的体验问题、针对性的满足差异化的体验诉求，让用户在使用产品时体验更佳，是其中的关键。

▶ 什么是用户体验?

突发异常

量级波动明显，容易监控



软件异常



团购无法使用

小音量异常

隐蔽性强，偶尔发生，问题严重



严重的底线问题

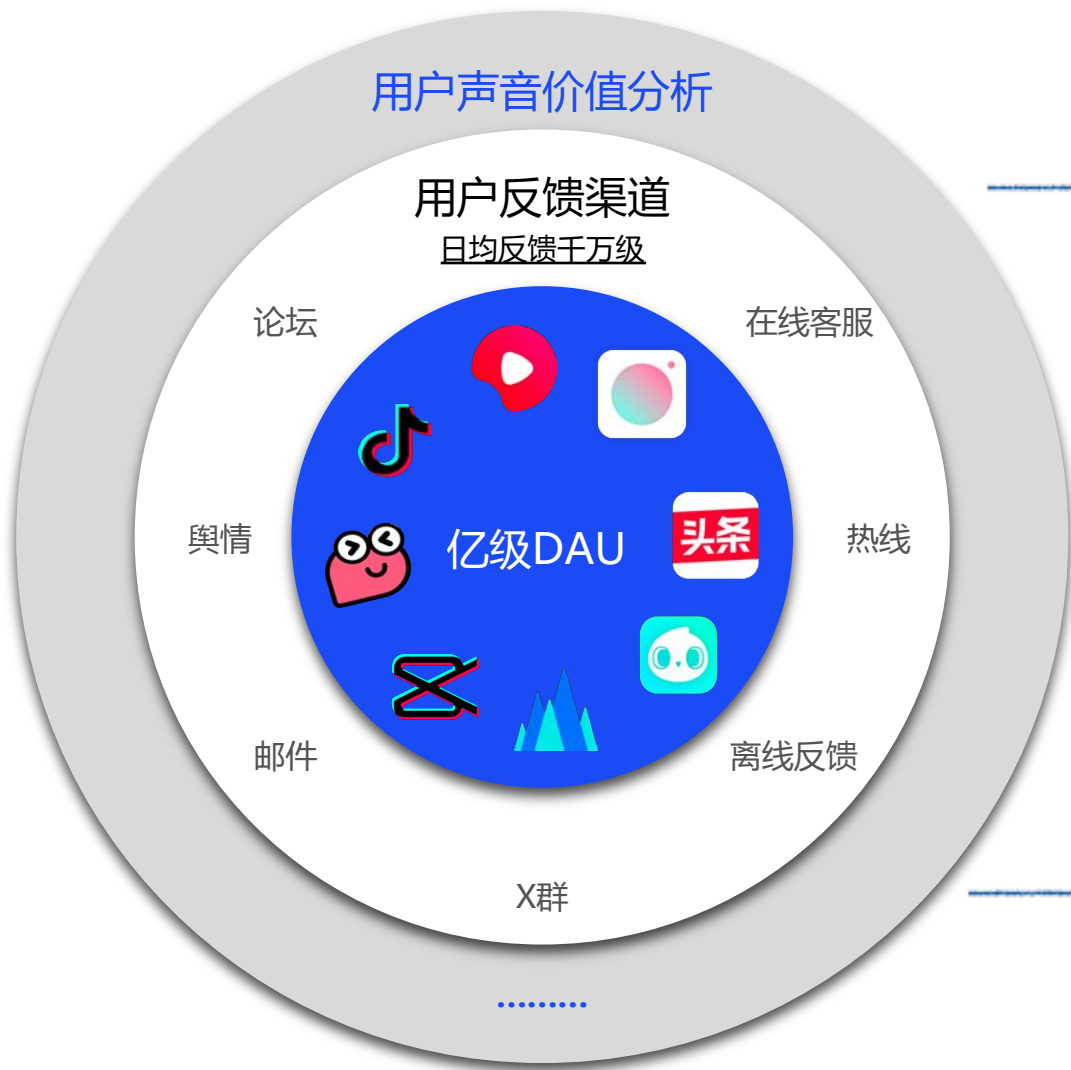
顽疾异常

隐含其它问题，难看清，长期存在

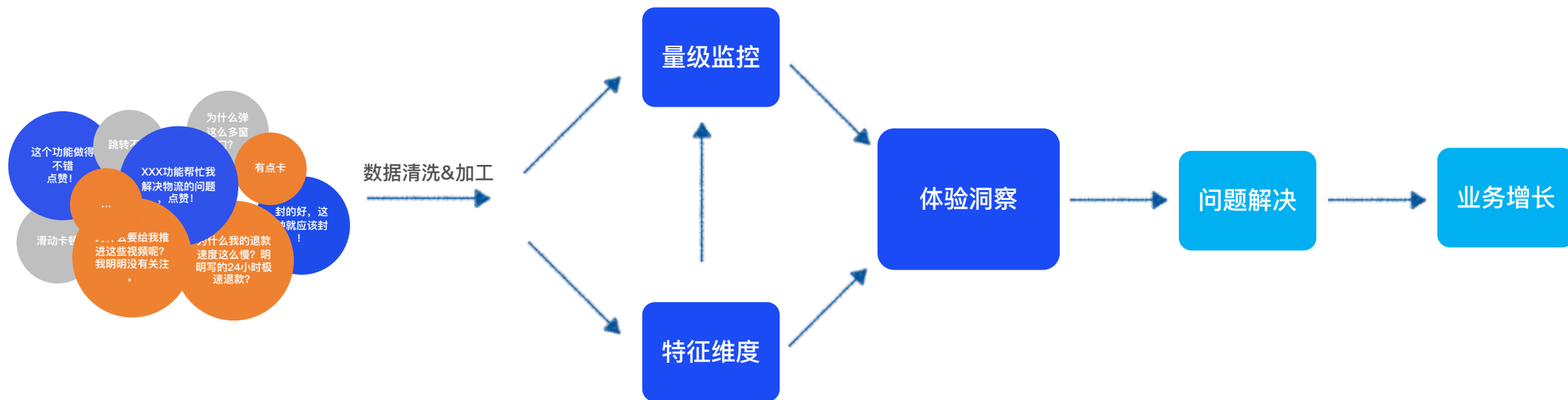


参加不了活动

▶ 用户体验中台是什么？



▶ 用户体验中台是什么？



从海量用户的声音中洞察体验问题，并推动问题解决，实现业务持续增长

PART 02

阶段一：引入AI工具来看清楚 体验问题

▶ 阶段一：如何从海量反馈中看清体验问题

反馈数据的多样性

40+反馈形式



问卷



在线客服



留言



论坛



电话



邮件

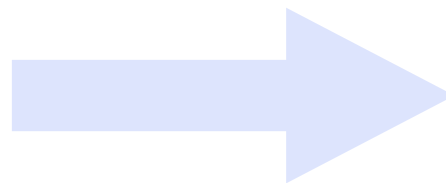
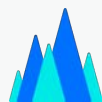
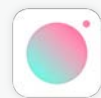


群聊



应用市场

亿级DAU的产品业务



体验问题的隐蔽性

推荐不喜欢

搜索不到

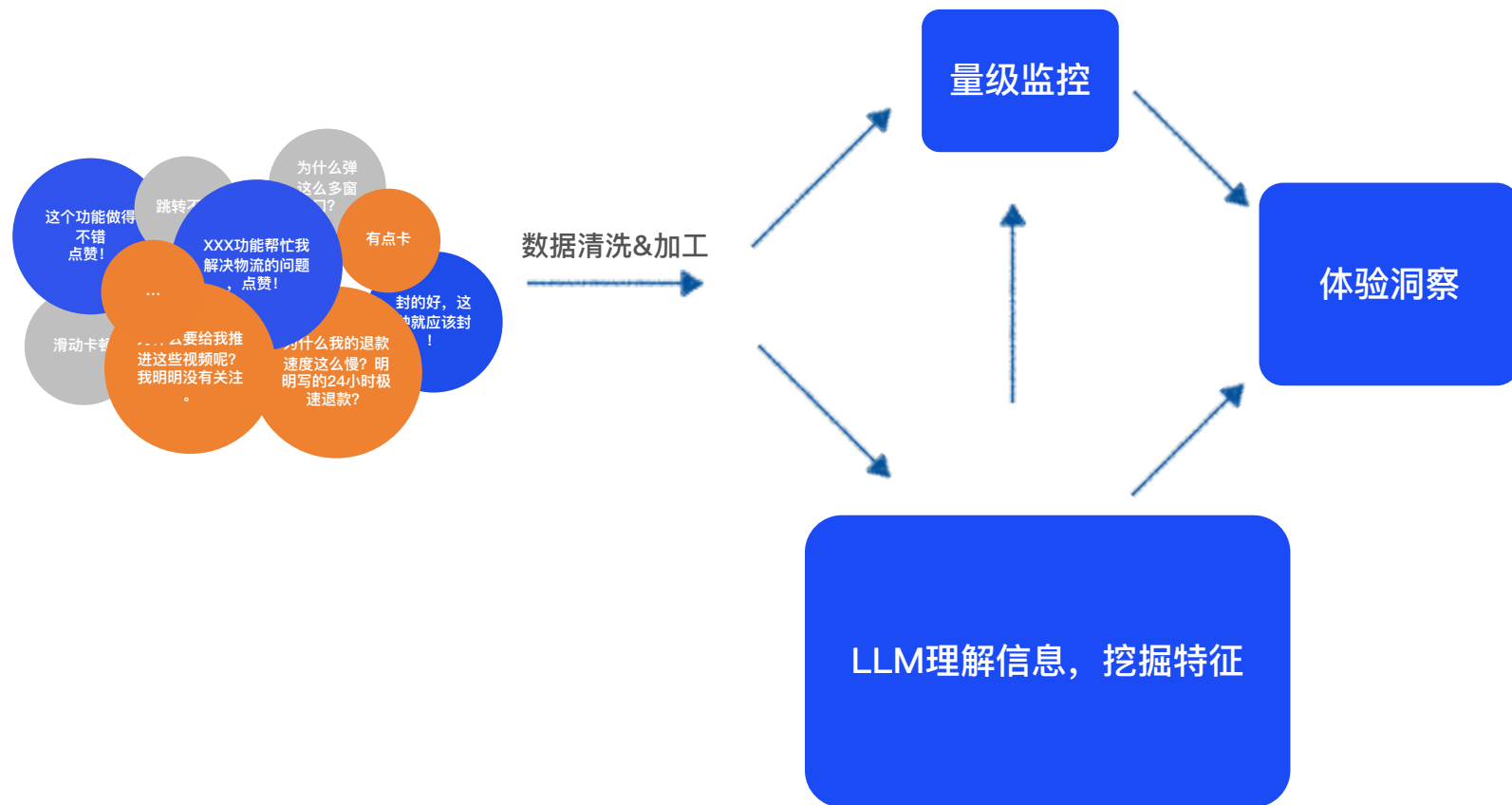
通知频繁

错误封禁

内容相似

更多体验问题

▶ 阶段一：通过AI工具来看清体验问题



▶ 看清用户体验问题 - LLM探索-摘要生成

为什么要摘要生成?

- 反馈内容冗长且阅读比较困难
- 问题分析时需要人工分析多个原声摘要, 十分耗时/费人力成本
- 原声内容知识点比较散碎, 不够凝练, 难以支持相关任务场景 (问题发现等)

什么样的摘要是好摘要?

- 摘出的问题是准确的
- 摘出的问题是完备的
- 摘出的问题**不存在编造**的问题

【用户】：申请退款
【客服】：您的订单已预约成功,目前无法申请退款!建议您按照预约时间前往体验哦
【用户】：还没有审核好吗?
【客服】：您好,请问您是否想咨询以下订单信息?「卡片消息」
【用户】：是的
【客服】：这边查看已有相关工作人员为您处理跟进了哈
【客服】：辛苦您耐心等待一下哈
【用户】：那请帮我反馈一下
【客服】：小助手帮您备注一下哈
... ..
【用户】：发票已经开好了
【用户】：叫他看一下
【客服】：好的,小助手帮你备注一下哈
【客服】：您好,请问是否仍需帮助,可以继续向小助手反馈

算法摘要：用户咨询订单退款审核进度并要求客服帮忙反馈。

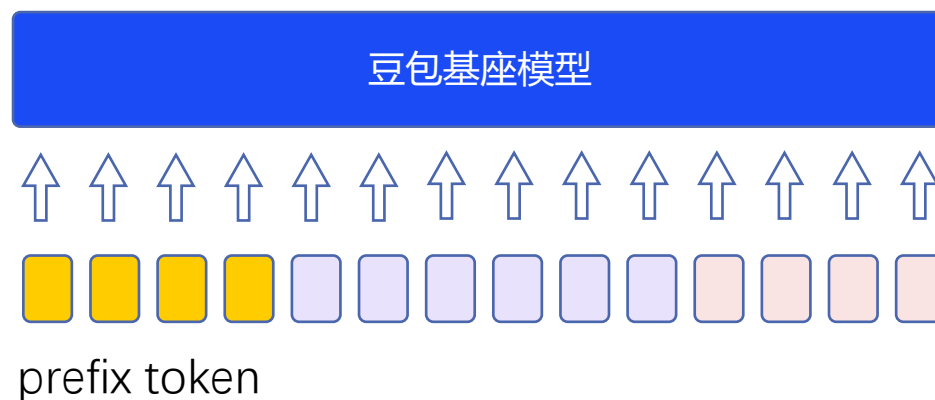
▶ 看清用户体验问题 - 解决LLM 的幻觉问题

PE 工程

约束:

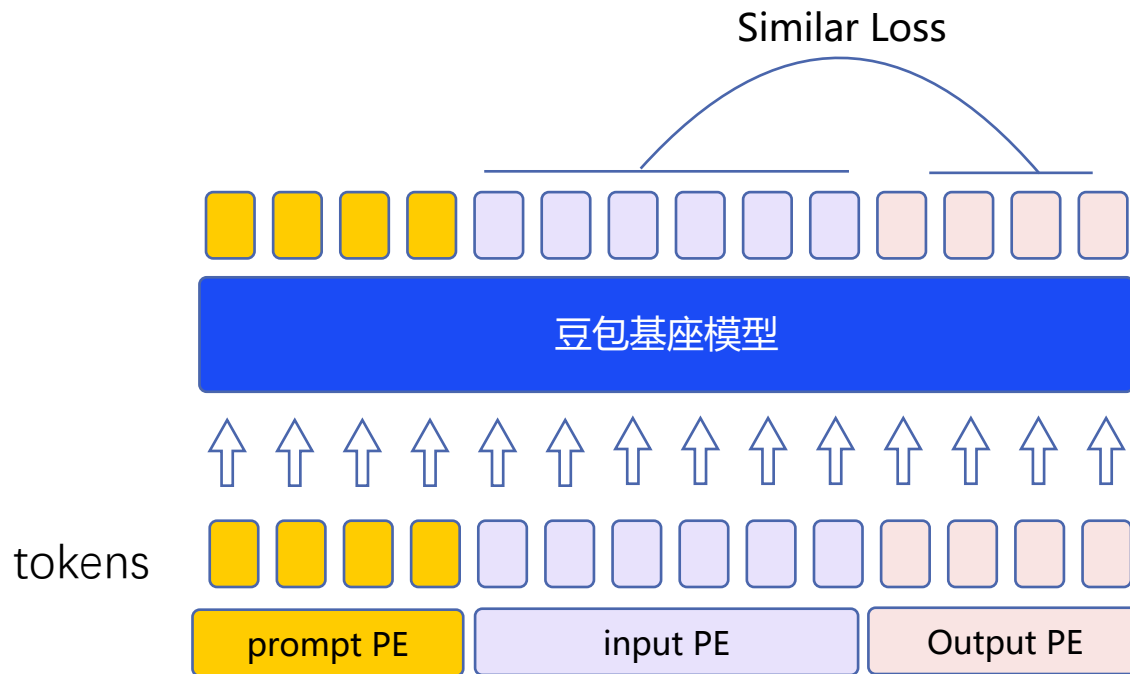
- 严格参考用户反馈内容进行摘要
- 确保摘要来源于反馈
- 禁止编造, 提供虚假内容
- 关注用户反馈的问题
- 忽略无关的特殊字符

P-tuning



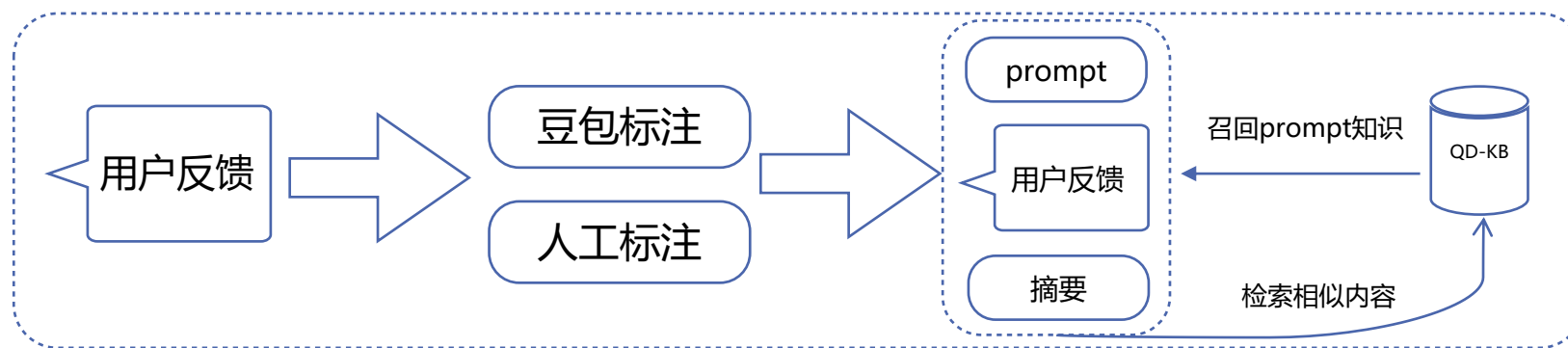
看清用户体验问题 - 摘要模型微调

模型微调

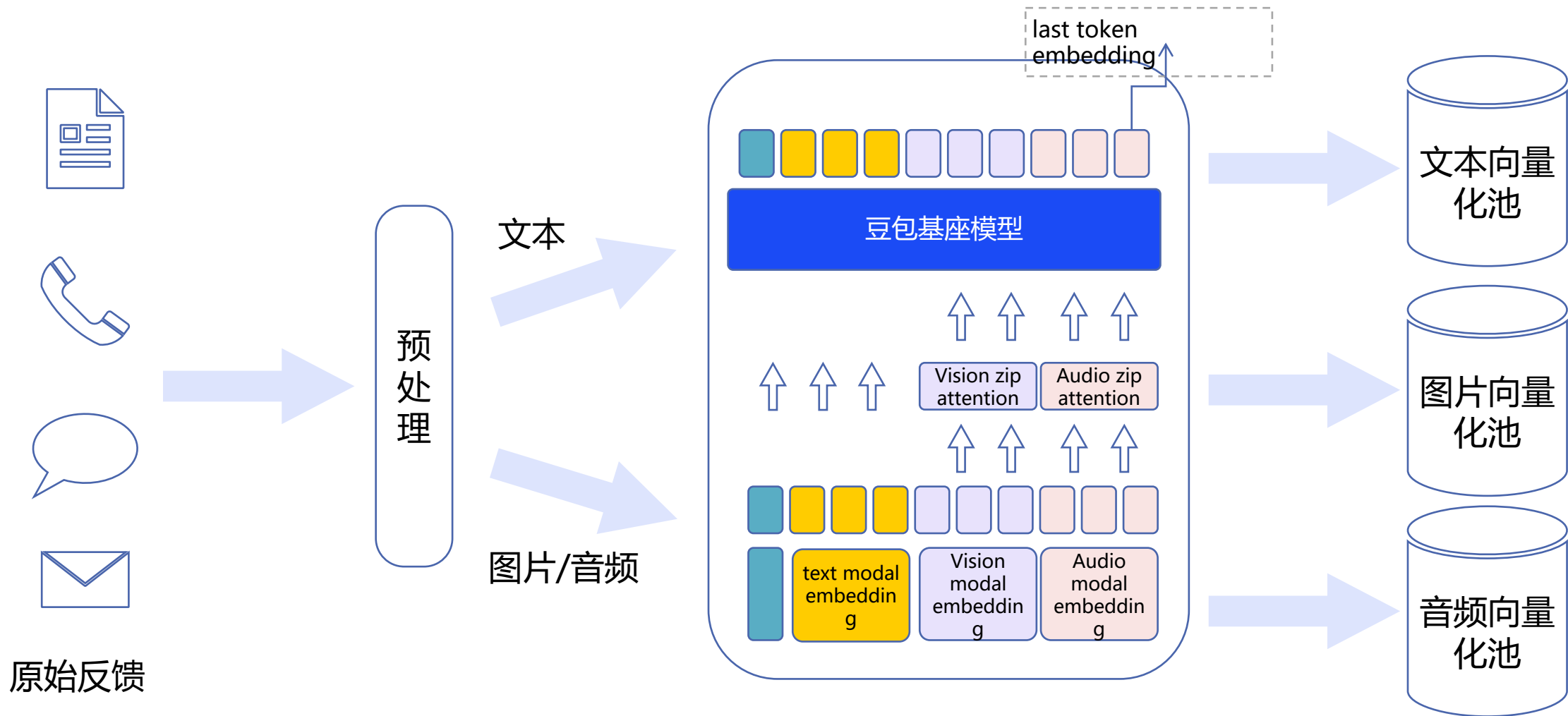


避免幻觉：
摘要与原文内容要足够相似

数据扩增

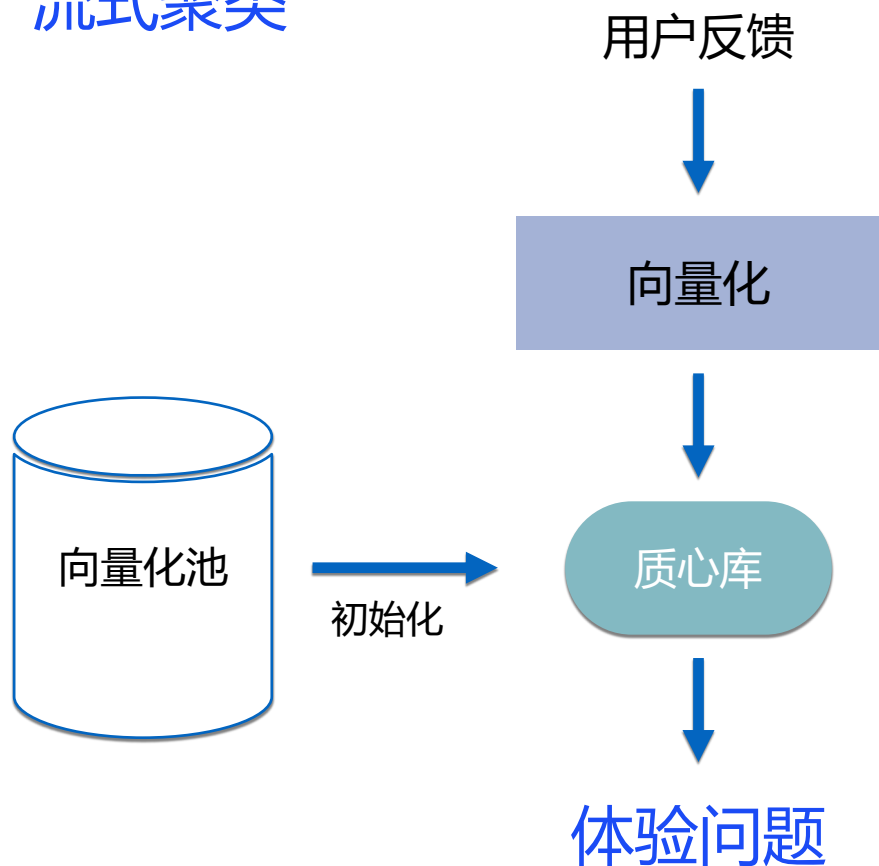


看清用户体验问题 - LLM构建反馈向量化池

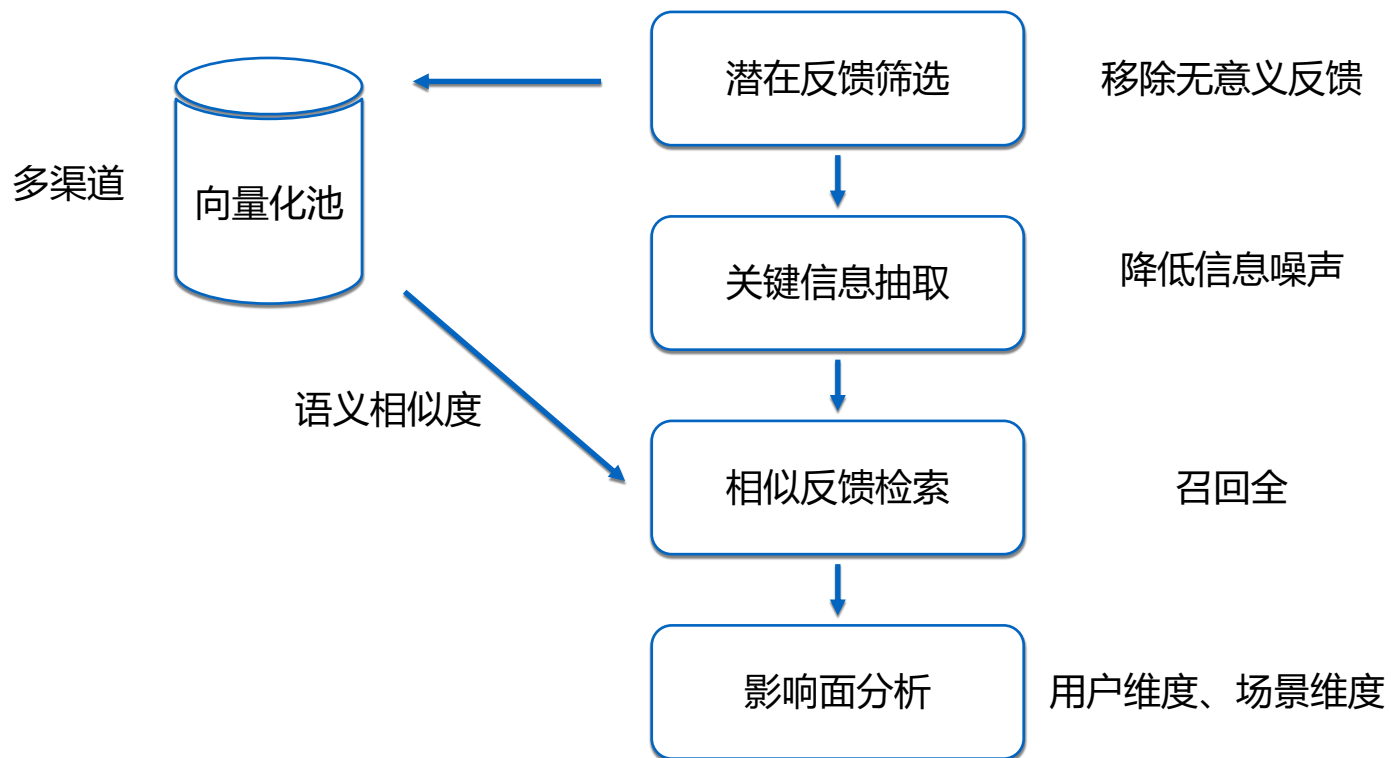


看清用户体验问题 - 向量化使用场景

流式聚类



相似问题召回评估



▶ 看清用户体验问题 - LLM进一步的使用场景

洞察体验问题

情感分析

摘要

批量总结

实体识别

文本标签

bug挖掘

LLM

情绪分布

挖掘到的情绪类型，关联分析

智能搜索

快速找到相关实体与标签，评估相关问题的反馈量级

相关实体问题

根据实体搜索到相关实体反馈声量

智能摘要/总结

快速阅读反馈内容，提升效率，批量总结反馈内容

异常识别

识别反馈中的bug问题，底线问题

▶ 阶段一：新的挑战

洞察体验问题

情绪分布

挖掘到的情绪类型，关联分析

智能搜索

快速找到相关实体与标签，评估相关问题的反馈量级

相似问题

根据语义相关性关联相似反馈

智能摘要/总结

快速阅读反馈内容，提升效率，批量总结反馈内容

异常识别

识别反馈中的bug问题，底线问题

新的挑战

如何更高效地支持
海量反馈X个性化场景

日千万级反馈

个性化场景

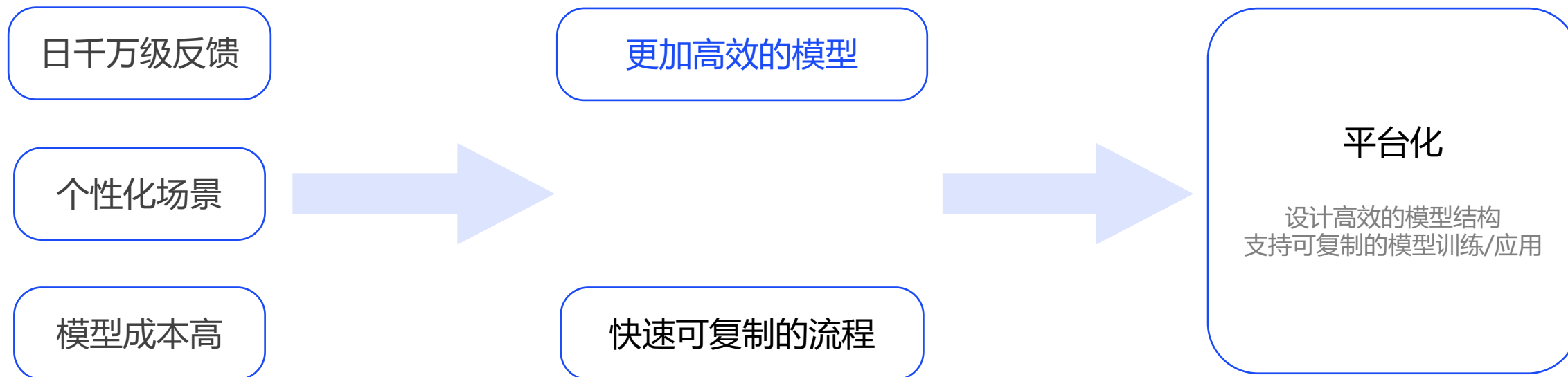
模型成本高

PART 04

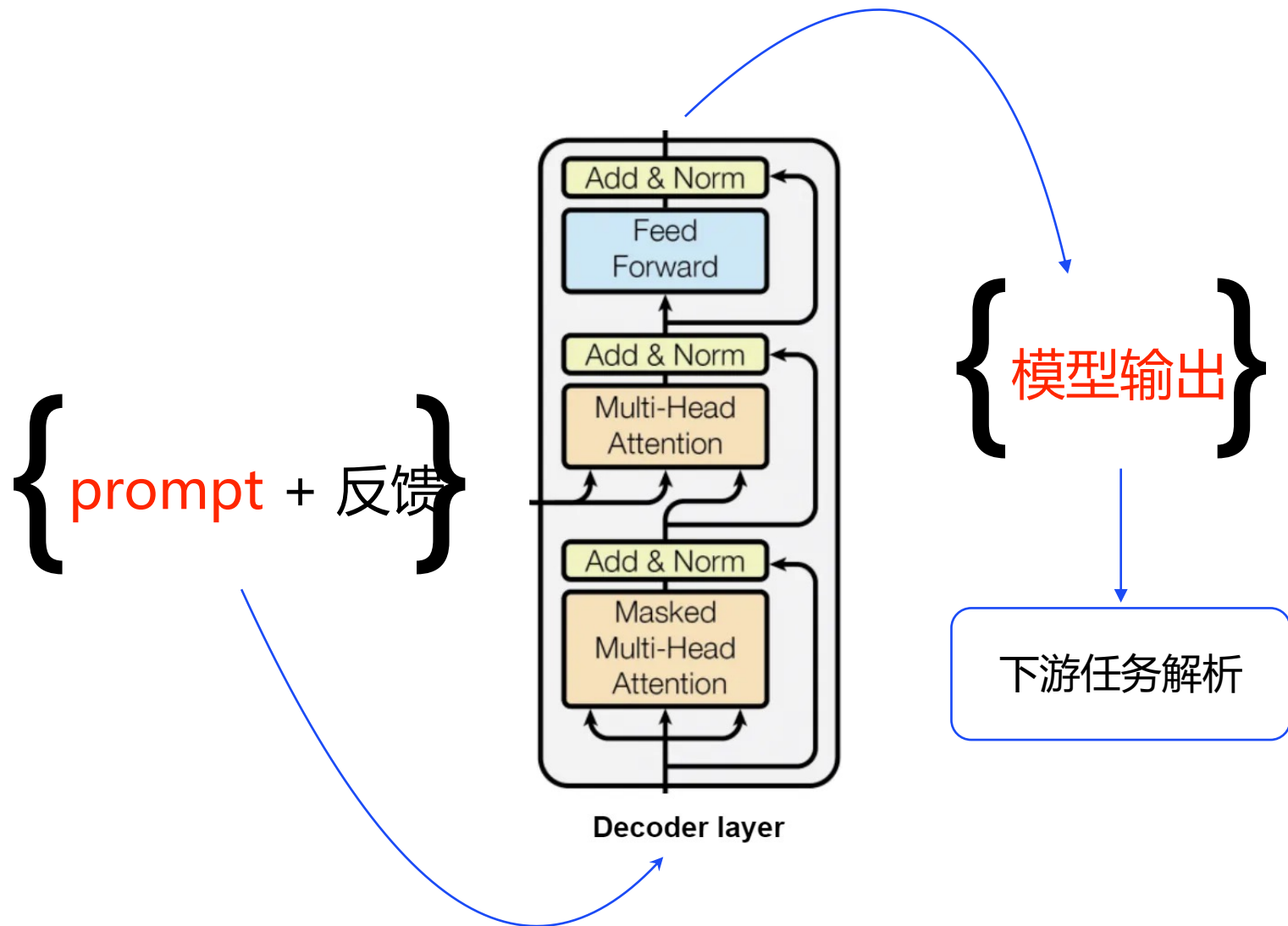
阶段二：大模型的工具化到平台化

▶ 阶段二：如何用好大模型

大模型应用下的新挑战



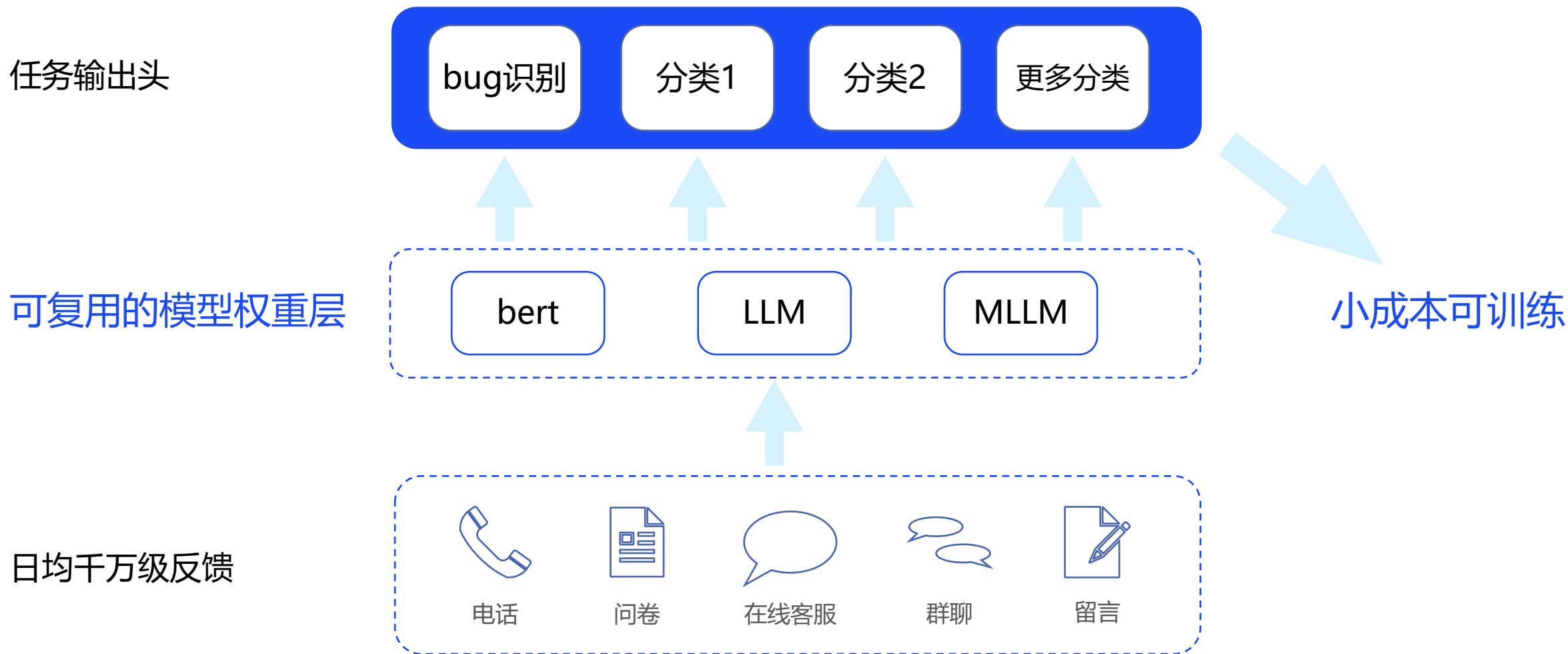
▶ 阶段二：如何用好大模型



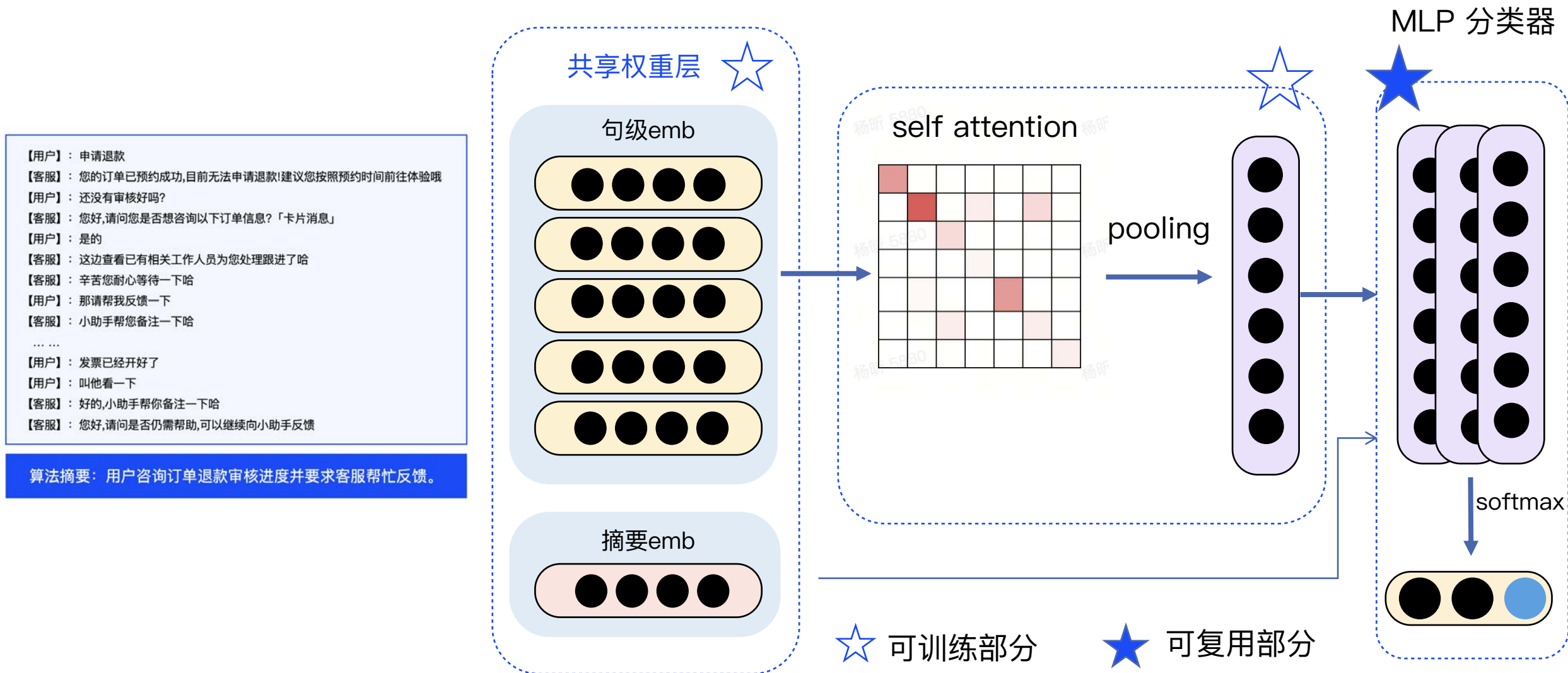
大模型成本高

不同任务难复用，输入的prompt和输出都有变化

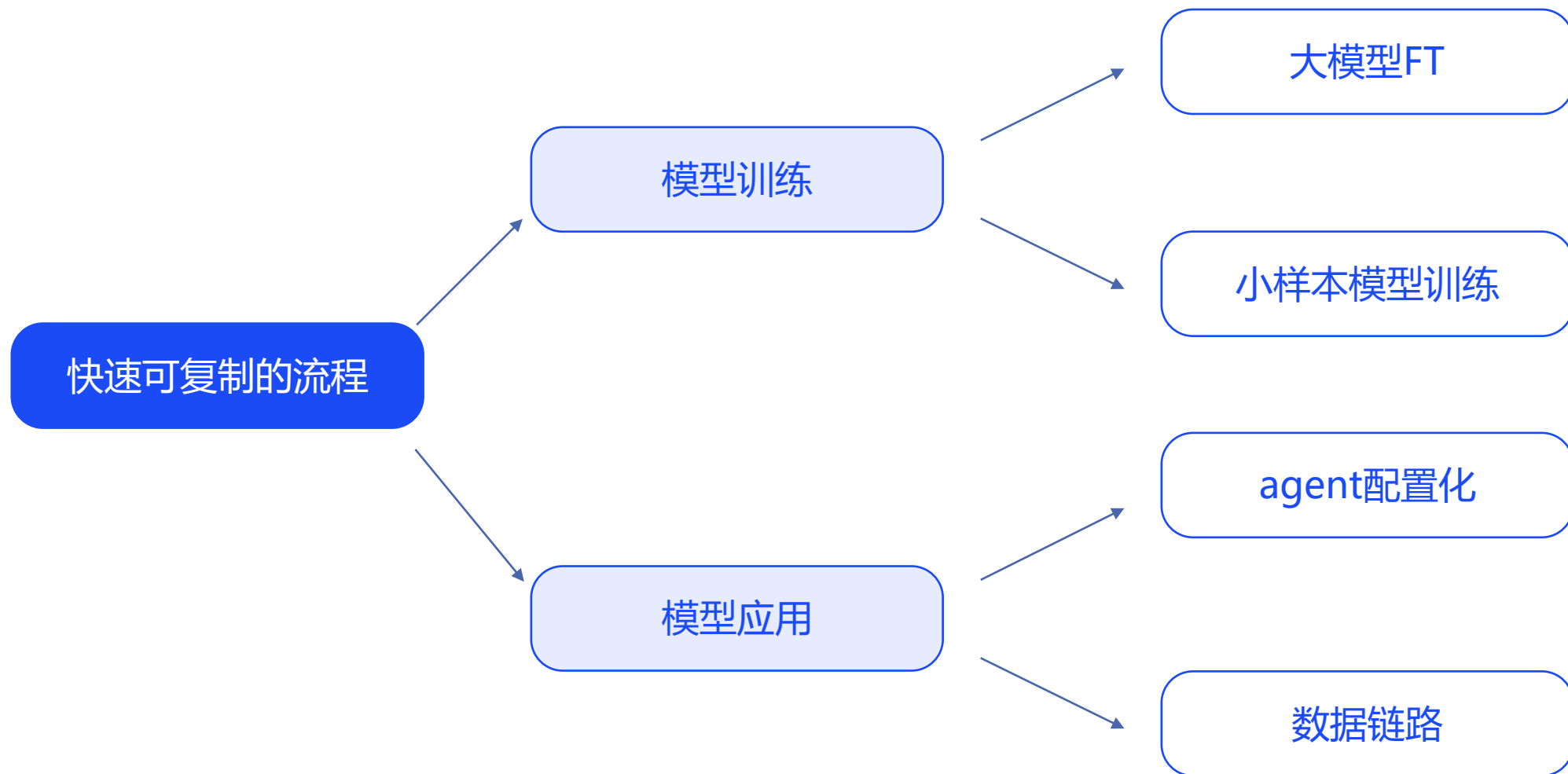
▶ 阶段二：思路 - 分离底座模型与分类任务



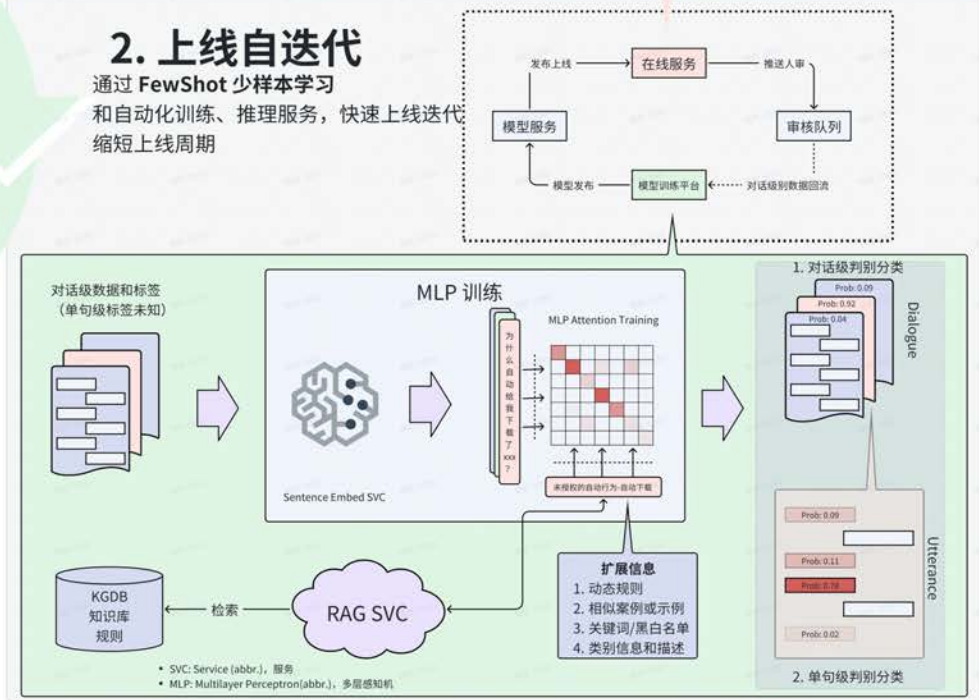
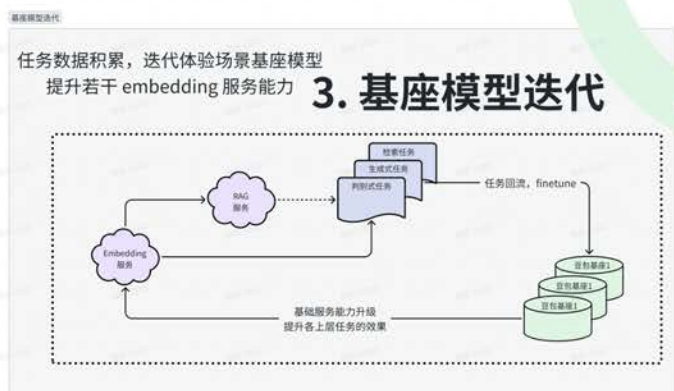
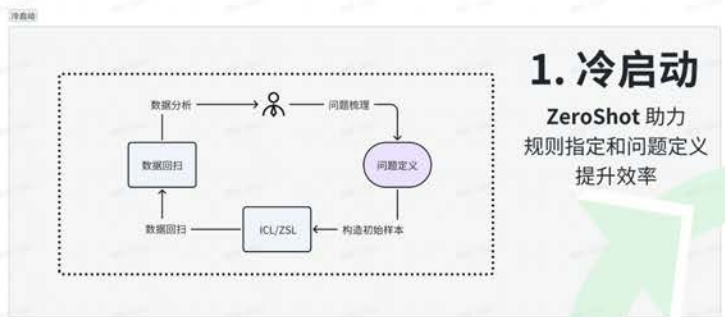
▶ 阶段二：小样本训练 - 解决大模型成本和个性化的问题



▶ 阶段二：如何用好大模型



阶段二：如何用好大模型 - 模型训练



冷启动阶段

少量数据
数据扩增
人工标注

小样本模型自迭代

少量数据
MLP训练
积攒数据

基座模型迭代

积攒大量数据
提升embedding

▶ 阶段二：模型应用 - agent配置化

体验应用

反馈摘要

异常挖掘

聚类告警

更多应用

Agent 平台

应用广场

租户内的Agent/Agent Flow

评估发布

评估

发布

调用日志

评估任务

标注

准召

发布范围

所属模块

内容保护

正向部署

数据统计

明细详情

关键能力

agent

flow

模型训练

prompt

api

agent节点

条件判断

SFT微调

任务调度

评估/评测

记忆变量

rag知识库

API节点

开始/结束

embedding

数据集

模型上线

基础设施

模型

账号体系

数据

基座LLM

基座MLLM

租户隔离

电话

留言

工单

问卷

.....

小样本模型

方舟模型

登录注册

阶段二：模型训练与agent应用整体架构

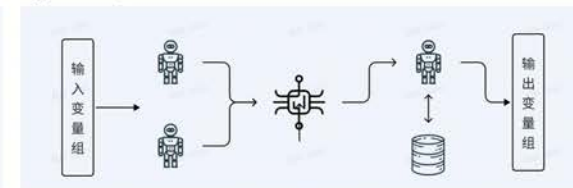
智能agent



非智能agent

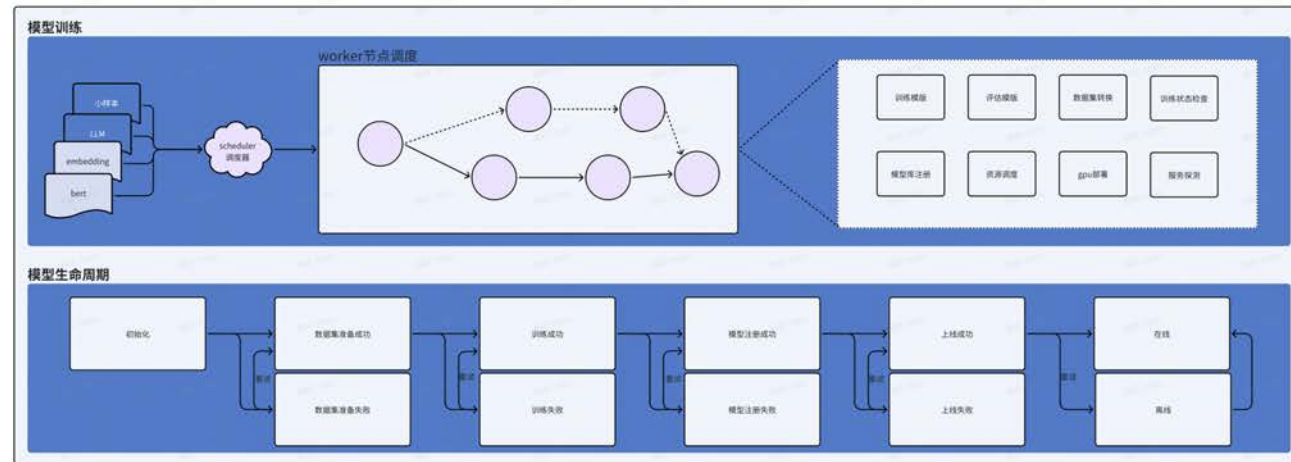
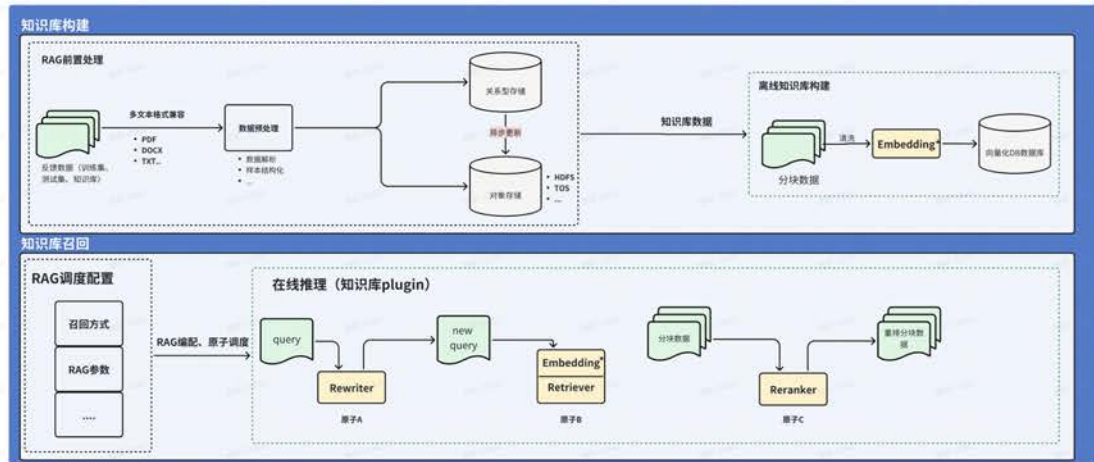
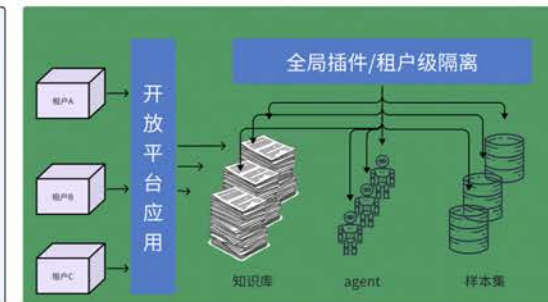
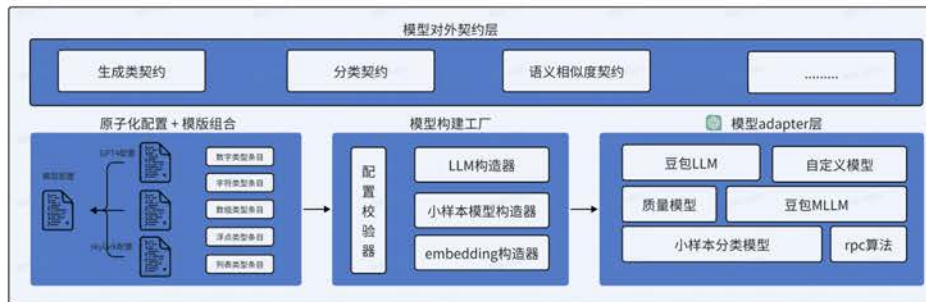


agent flow/工作流编排

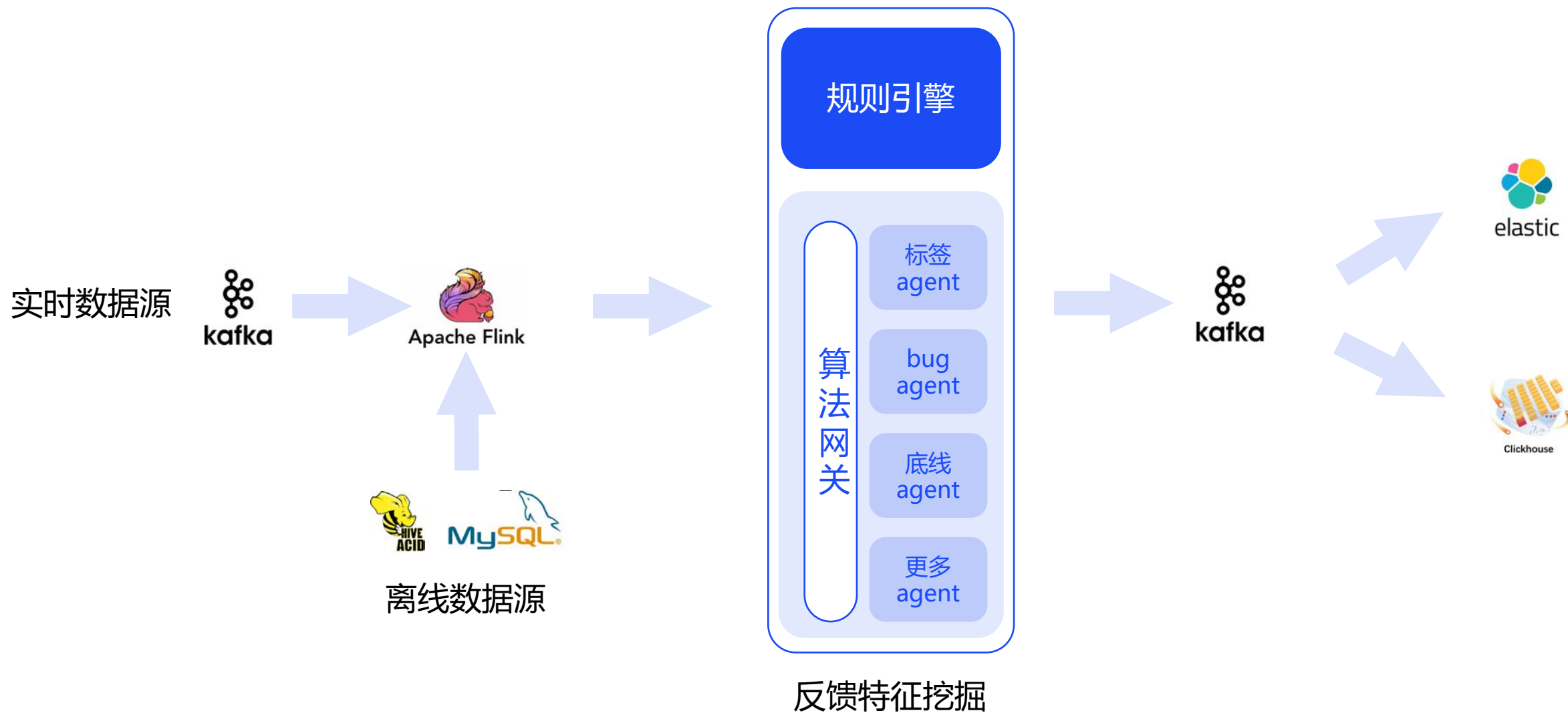


全程上下文
共享环境变量

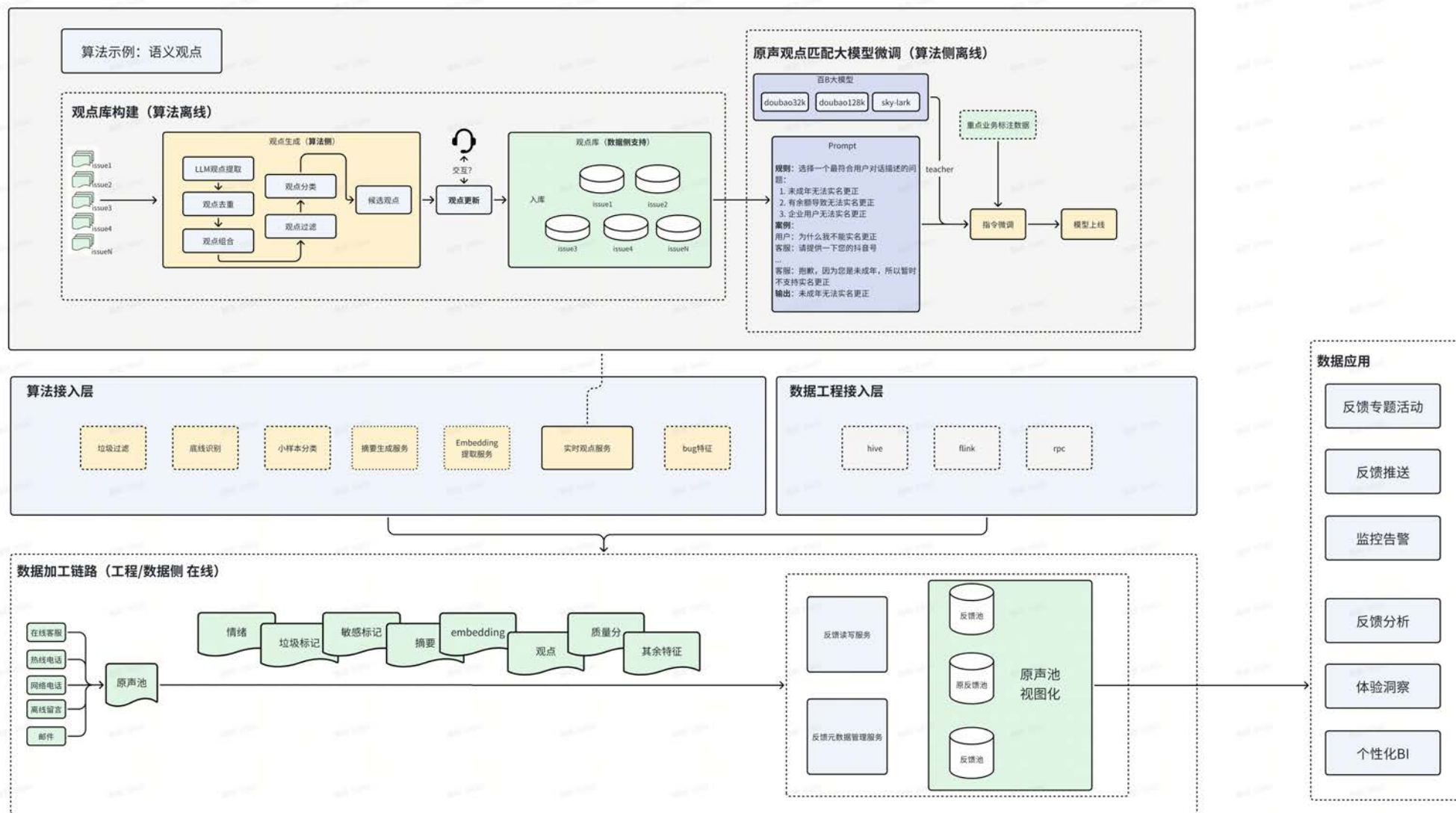
反馈数据源



▶ 阶段二：模型应用-可插拔的AI数据链路



阶段二：模型应用-数据链路整体架构图



▶ 阶段二：效果

小音量异常(召回XX+)



严重的底线问题

顽疾异常(召回XXXX+)



实名认证失败

覆盖
千亿级反馈和xxx+个性化场景

召回体验问题XXXX

P0/P1严重体验问题XXX

体验问题解决率 XX%

百万DAU用户反馈求助率下降 2x%+

PART 04

回顾与总结

▶ 阶段一：如何从海量反馈中看清体验问题

反馈数据的多样性

40+反馈形式



问卷



在线客服



留言



论坛



电话



邮件

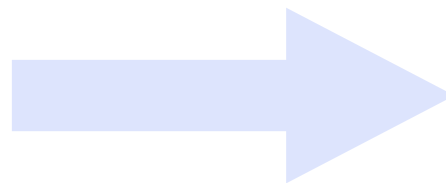
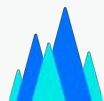


群聊



应用市场

亿级DAU的产品业务



体验问题的隐蔽性

推荐不喜欢

搜索不到

通知频繁

错误封禁

内容相似

更多体验问题

▶ 阶段一：引入AI工具来看清体验问题

文本

文本分类

文本聚类

实体识别

语义相似

摘要

批量总结

情感分析

文本标签

bug挖掘

bert

LLM

图片/音频/视频

OCR

相似度

音频识别

实体识别

Speech-to-Text

CNN

OCR

MLLM

▶ 看清用户体验问题 - LLM的新挑战

实体识别

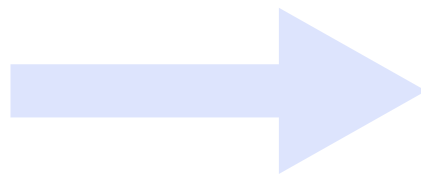
摘要

批量总结

情感分析

文本标签

bug挖掘



日千万级反馈

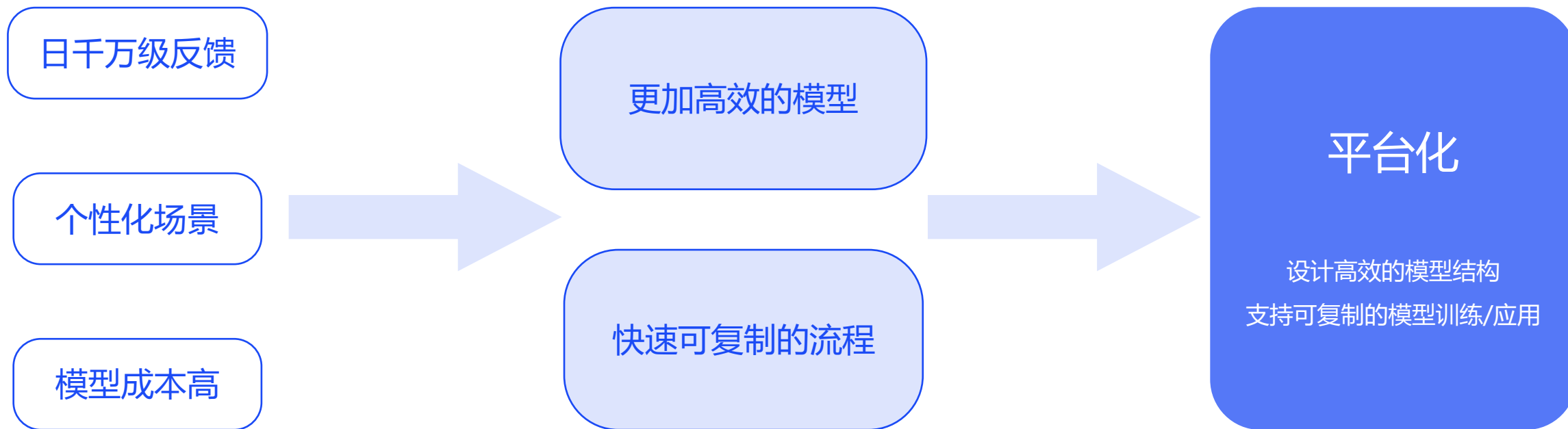
个性化场景

LLM使用场景广泛

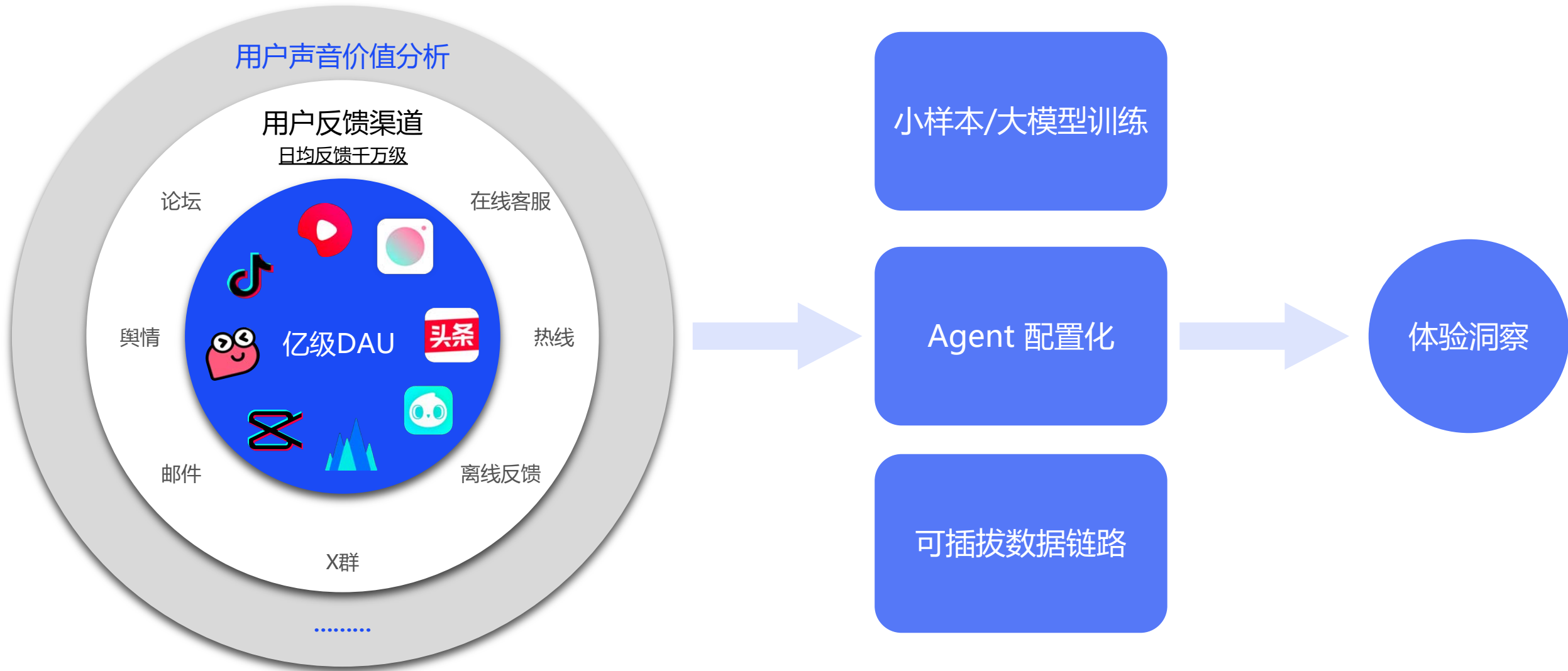
模型成本高难以支撑

▶ 阶段二：如何用好大模型

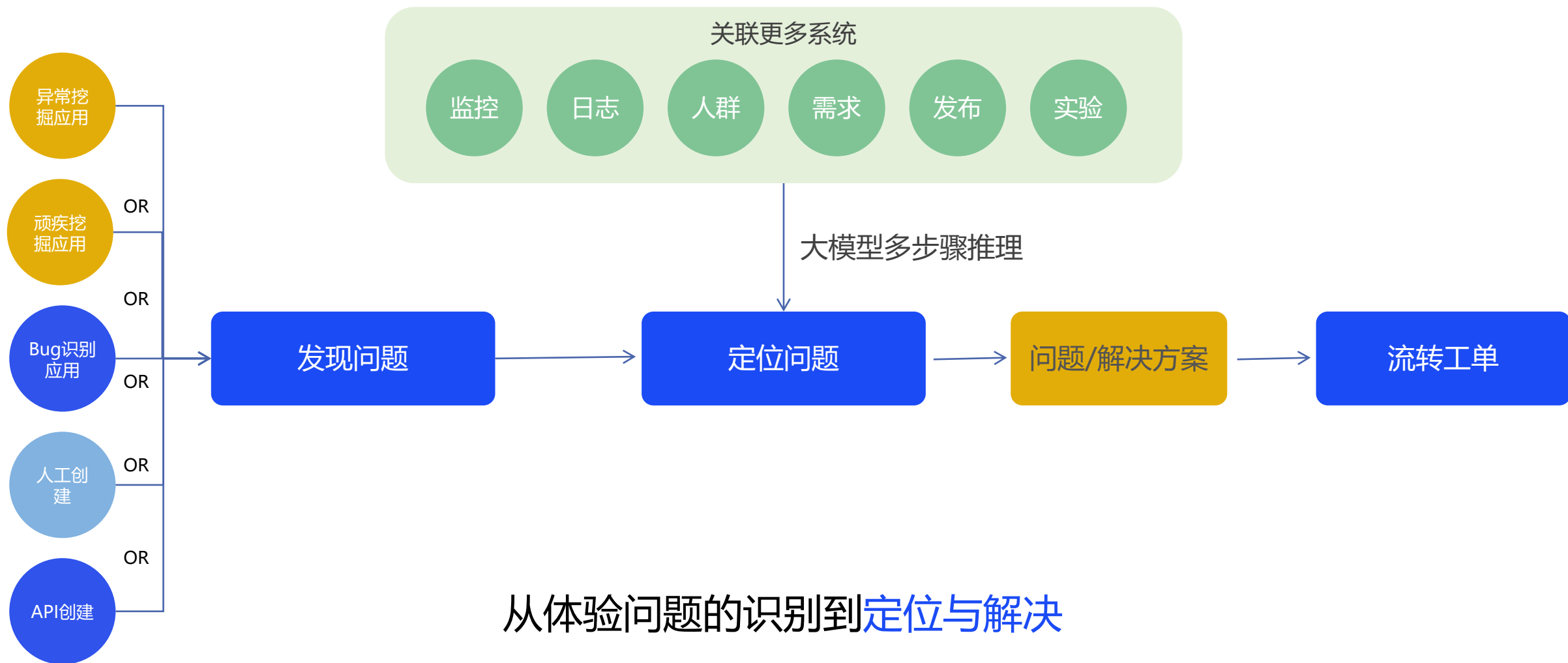
从点到面：大模型从工具化到平台化



总结与展望 - 通过大模型平台化建设用户体验中台



总结与展望 - 下一个阶段



科技生态圈峰会 + 深度研习



—1000+ 技术团队的选择



 **K+峰会**  **敦煌站**

K+ 思考周®研习社

时间: 2025.08.29-30

 **K+峰会**  **上海站**

K+ 金融专场

时间: 2025.10.17-18

 **K+峰会**  **香港站**

K+ 思考周®研习社

时间: 2025.11.25-26



K+峰会详情



 **AiDD峰会**  **上海站**

AI+研发数字峰会

时间: 2025.05.17-18

 **AiDD峰会**  **北京站**

AI+研发数字峰会

时间: 2025.08.08-09

 **AiDD峰会**  **深圳站**

AI+研发数字峰会

时间: 2025.11.28-29



AiDD峰会详情



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时代

