

QCon  
全球软件开发大会

# 多智能体驱动的企业级 ChatBI 落地实践

陈叶超

喜马拉雅数据平台负责人



# 目录

01 喜马 ChatBI 背景和挑战

02 喜马 ChatBI 产品与架构

03 构建可持续进化 ChatBI

04 未来展望和规划

📍 北京

**QCon**

全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

📍 北京

**AiCon**

全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

📍 上海

**QCon**

全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月

5月

6月

8月

10月

12月

📍 上海

**AiCon**

全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

📍 深圳

**AiCon**

全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

📍 北京

**AiCon**

全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOPs
- 具身智能

# 极客邦科技 2025 年会议规划

促进软件开发及相关领域知识与创新的传播



参会咨询



查看会议

01

# 喜马拉雅 ChatBI 背景和挑战

# ChatBI 研发背景



用数门槛高：找数难、取数难、不会写 SQL  
数据不平权：业务团队排队等需求，需求响应慢、依赖数据产品和自助取数

## 业务痛点



数据消费效率低、数据价值得不到释放  
人力成本高、交付压力大  
需求被动、数据需求投诉后才知道  
重复取数，成就感低

## 数据团队痛点



大模型和智能体技术发展迭代快，能力进度显著

传统 BI 加速向 BI+AI 转型

## Data+AI 发展趋势



自然语言是最高效、最直接、最有价值需求表达和反馈

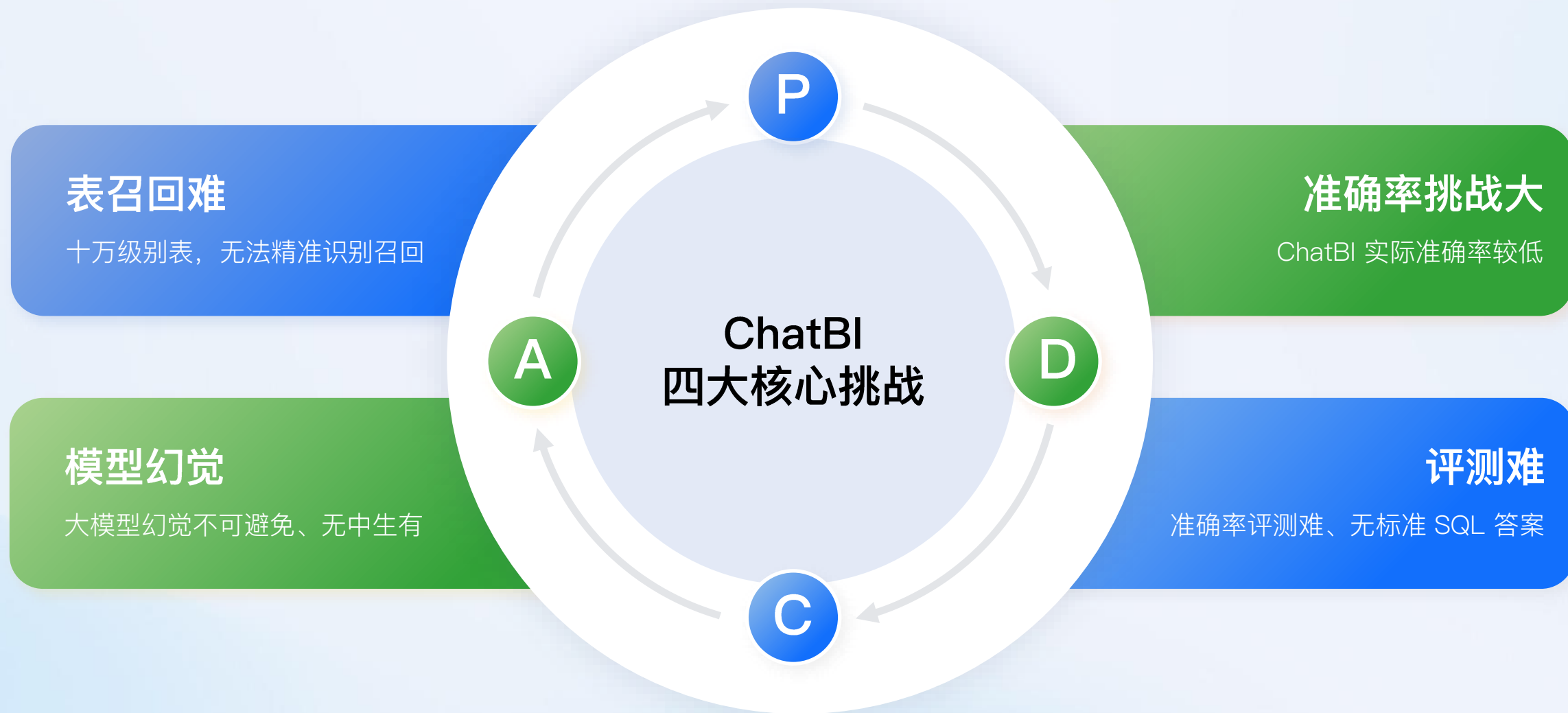
## 我们的观点

# 为什么要自研

喜马 ChatBI 与 ChatBI 商业化产品的核心差异

| 项目                | 开源&商业化产品                         | 自研喜马 ChatBI                                        |
|-------------------|----------------------------------|----------------------------------------------------|
| 准确度               | 测试下来没有符合预期产品                     | 数百列+多表：至少85%以上，且能持续进步                              |
| 数据安全              | 无法继承原有数据权限、不支持加密、脱敏<br>安全合规存在挑战  | 复用现有数据权限、支持敏感加密、脱敏<br>安全合规                         |
| 智能选择数据<br>&多表关联   | 智能选择数据集基本不支持<br>多表关联基本不支持、效果比较差  | 自然语言提问情况下自动选择数据集<br>支持多表关联                         |
| 多轮问答/追问           | 有限支持                             | 支持                                                 |
| 开放、兼容             | 开放性差<br>模型、数据源、BI产品绑定            | 自身智能体扩展开放<br>向公司 Agent 应用开放<br>复用兼容当前 Lakehouse 架构 |
| 能力可进化<br>全链路追踪&反馈 | 有限预设支持<br>多数依赖模型微调、周期长效果差<br>N/A | 可调优、可介入、且快速、实时<br>查询速度、资源消耗、多维度的准确率分析等             |

# ChatBI 研发挑战



02

# 喜马拉雅 ChatBI 产品与架构





# ChatBI 产品架构

## 产品形态

ChatBI 桌面+移动端

ChatBI 办公工作台

ChatBI MCP Server

ChatBI OpenAPI

## 产品能力

智能选数据集

智能问数

单表问答

联表查询

多轮问答

智能绘图

取数解释

智能推荐

数据集管理

## 智能体能力

意图识别  
agent

智能选数据集  
agent

智能改写  
agent

SQL  
agent

智能问数  
agent

智能评测  
agent

用户理解  
agent

## Tools

RAG

Memory

DB查询

语法检查

权限校验

反馈&日志追踪

数据下载

取数过程查询

数据标注

## Context 增强

表/字段/值

业务知识

专业词汇

时间规则

指标知识

样例学习

SQL方言学习

个性化Memory

## 模型接入

模型接入管理

Embedding 模型

文本模型

...

## Data Infra

Milvus

StarRocks

Lakehouse

OneMeta

DataOps 平台

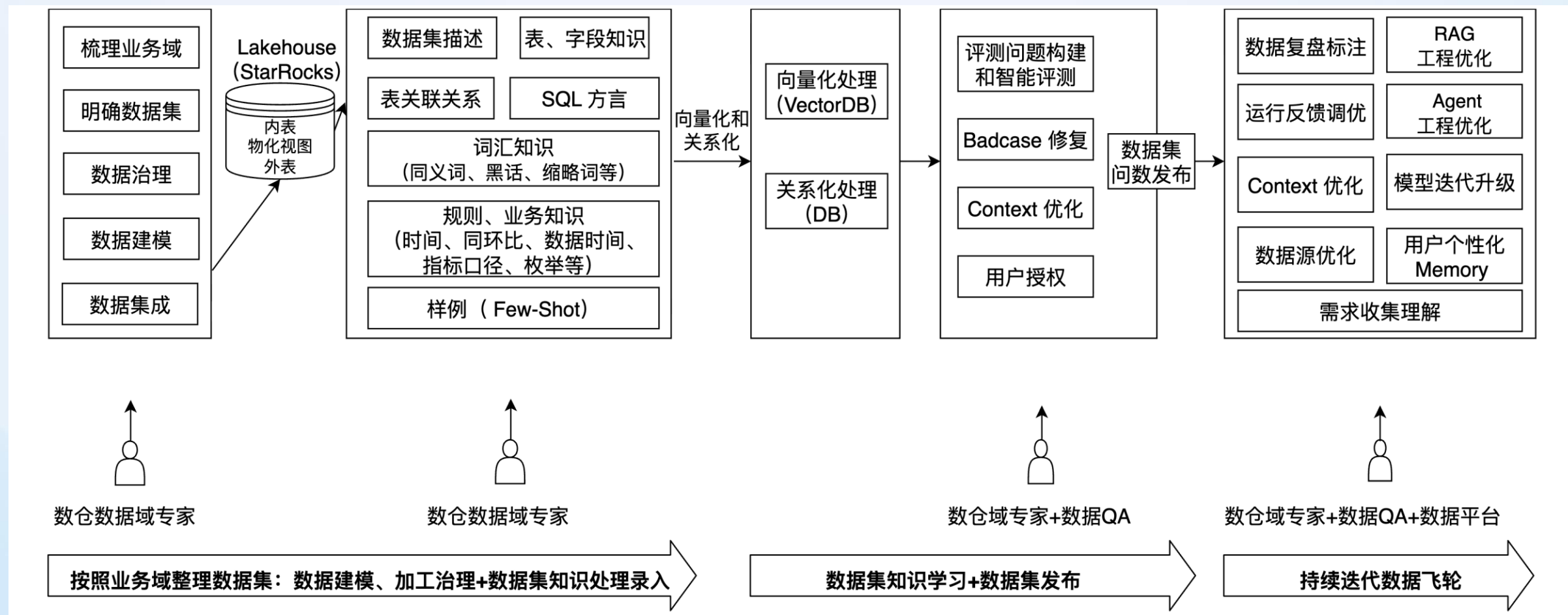


# 03

## 构建可持续进化的 ChatBI



# ChatBI 全生命周期



# 构建可持续进化的ChatBI

## ChatBI 四大要素

### 评测反馈机制

评测先行  
自动化可衡量的评测机制  
AI/人工标注

### 多 Agent 协同

按照核心职责拆分 Agent  
Agent少而精，每个Agent做到极致  
选择合适的LLM，且持续迭代



### 高质量数据集

按照业务域设计的高质量数据集  
精挑细选经过数据治理过的数据表  
精心维护的知识、规则、指标、语法等

### 个性化 Memory和用户理解

短记忆：会员上下文理解，追问等

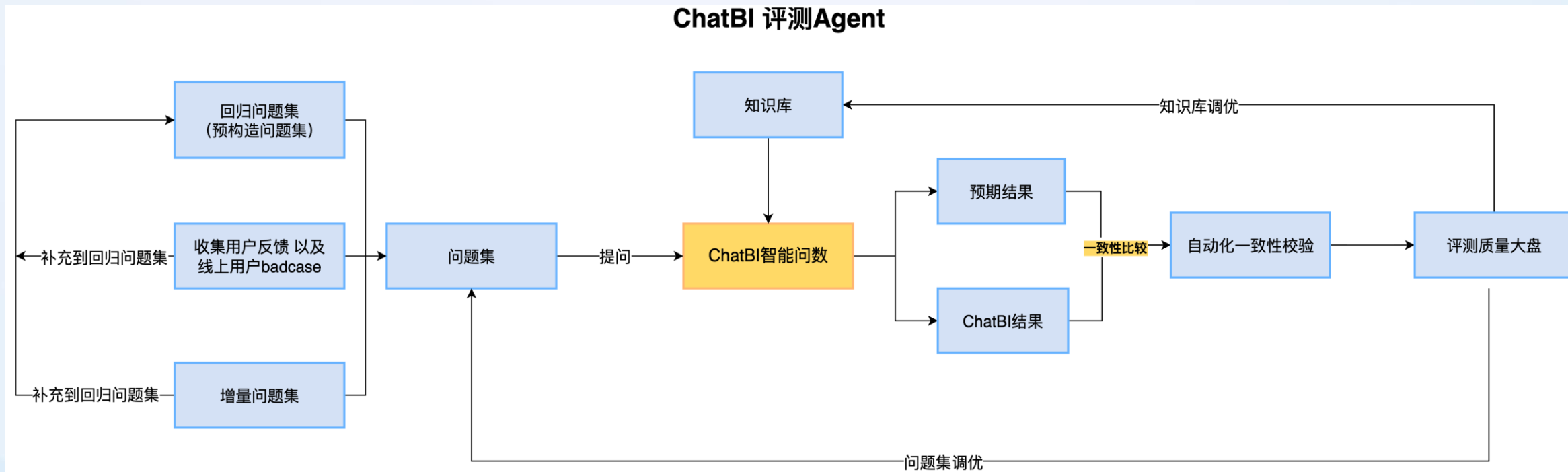
长期记忆：历史用户提供的知识、用户偏好推荐、个性化，冷启动/个性化推荐更精准

用户理解和需求引擎

# ● 评测难：无法衡量，就无法改进

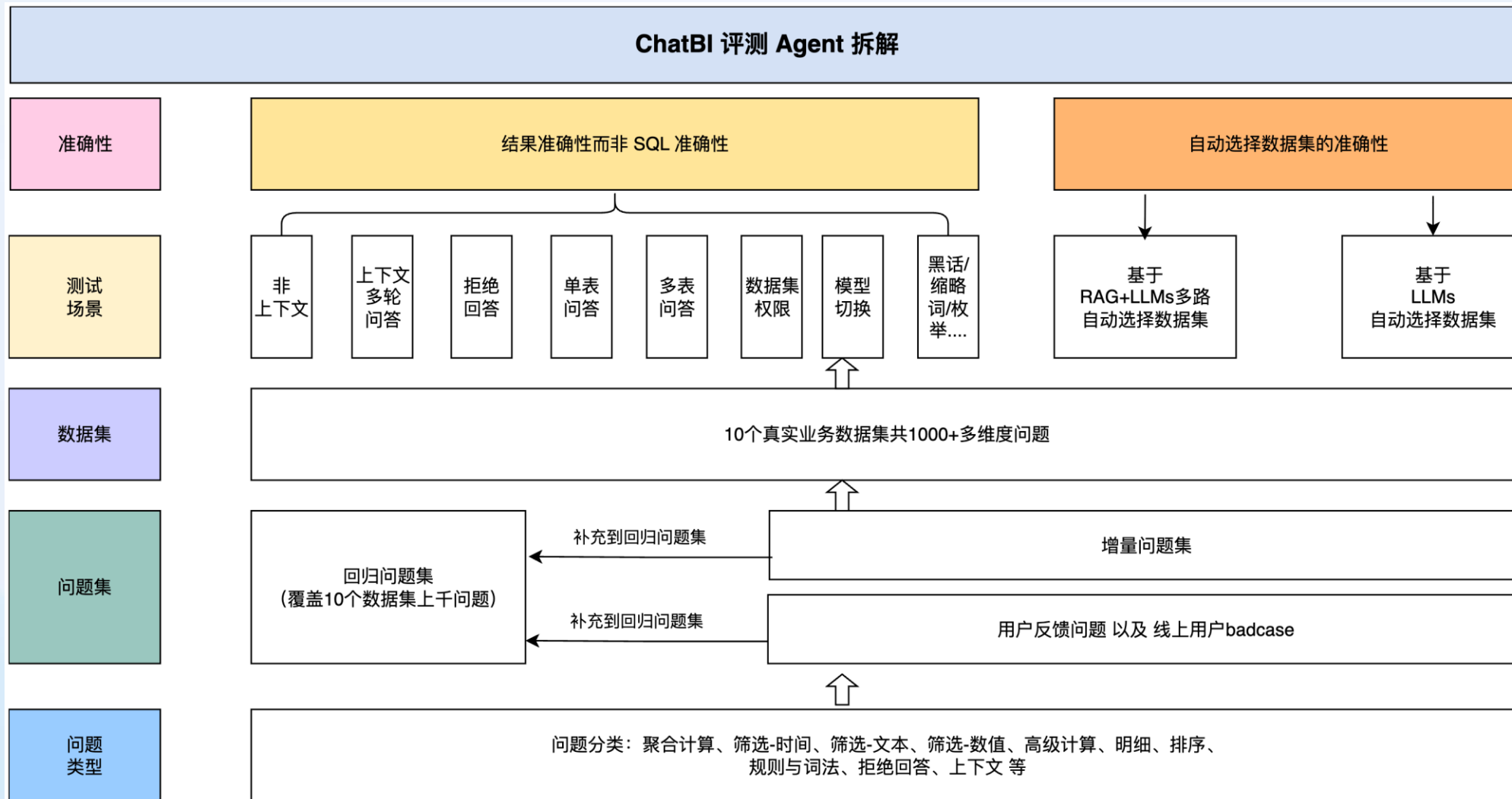


# 评测先行：无法衡量，就无法改进





# 评测先行：无法衡量，就无法改进



# 海量表召回的挑战



# 海量表召回的挑战



# ● 高质量数据集：解决召回难题

高质量数据集：对表、知识、规则等核心 Context 的抽象封装

表信息

表信息

字段信息

字段枚举

数据时间

是否维度...

知识

企业知识

黑话/术语

业务说明

指标口径

词汇说明...

规则

语法规则

时间规则

安全规则

Join规则...

其他

权限信息

预设推荐问题

样例

添加标题

添加标题

添加标题

# 高质量数据集：最佳实践



## 处理数据集而不是表

数据集是 ChatBI 处理 Context 工程最核心单位。  
召回数据集而不是表；基于数据集结构化信息生成SQL、处理权限



## 按业务域建设高质量数据集

大而全不如少而精  
业务域名负责人梳理核心业务表和知识库  
看板和自助取数重要性和访问量作为建设数据集重要参考



## 基于数据集优化知识

全局相关知识和单数据集相关知识  
对数据集知识进行高度数据治理和优化，经过评测以后再上线



## 软件工程概念仍然适用

数据集结构化处理，而不仅仅是向量化  
数据建模、关联度高的表打宽、OLAP 加速查询  
用好工具，比如复用数据集的权限、SQL 语法校验

## 高质量数据集：数据集实例

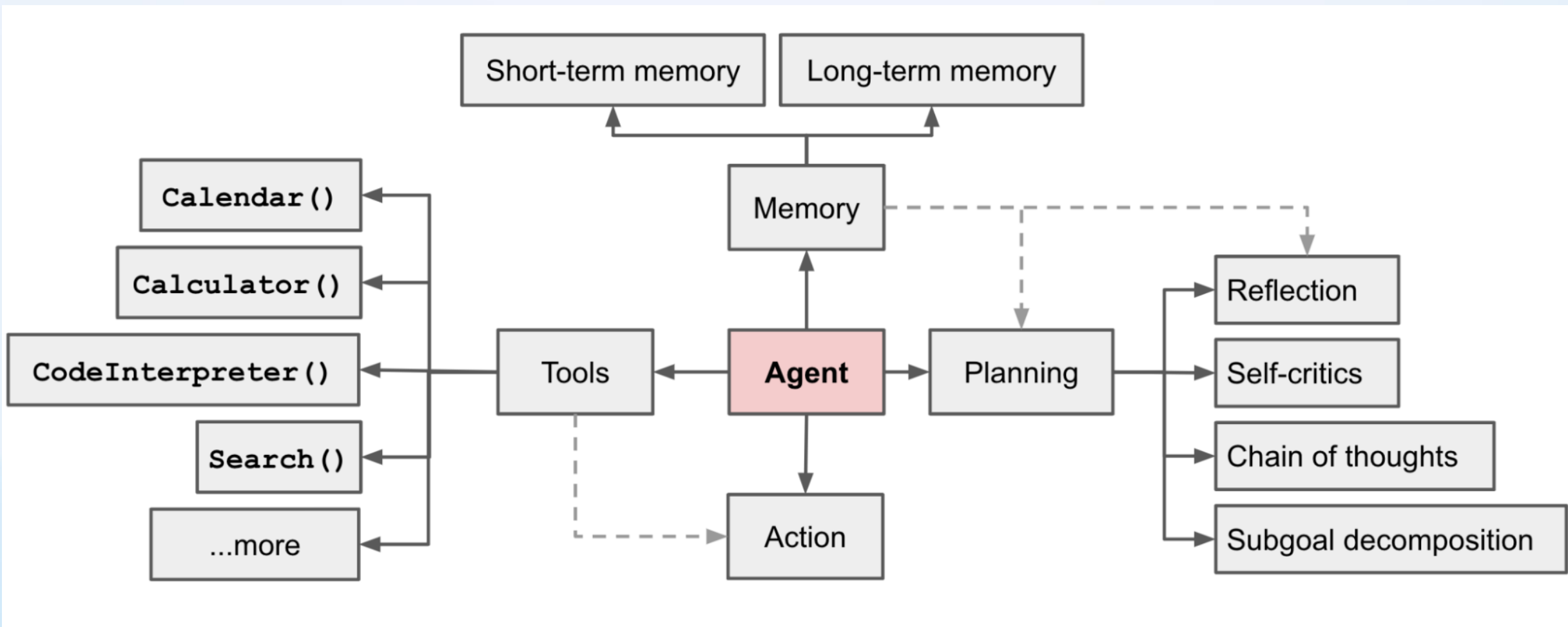
| 业务域   | 数据集     | 核心表                                       | 核心企业知识                             | 经典用户场景                                        |
|-------|---------|-------------------------------------------|------------------------------------|-----------------------------------------------|
| 公司经营域 | 收入和利润   | 整体经营收入<br>业务线收入等                          | 业务线专业词汇、黑话<br>业务线之间关系、指标口径         | 今年会员业务收入趋势如何                                  |
| 内容运营域 | 专辑综合分析集 | 专辑宽表、播放、<br>互动、流量归因等                      | 标签分类、赛道、频道、版权等                     | 最近一个月最热门专辑top 100，同环比如何<br>帮我圈100个热门科技财经类播客专辑 |
|       | 声音综合分析集 | 声音基本信息、播放、<br>互动指标等                       | 标签、是否上架、付费等枚举                      | 最近一周评论互动率最高的声音top10                           |
| 用户域   | 用户综合分析  | 用户/设备宽表、专辑<br>维度、画像等                      | 人群分类<br>互动、播放、留存指标口径、<br>黑话、资源位信息等 | 今年DAU 7日留存日均趋势<br>top10的资源位                   |
| 创作者域  | 主播分析    | 创作者基本信息<br>创作者专辑、声音、<br>直播统计信息<br>收入统计信息等 | 主播分级分类、赛道                          | 最近半年订阅订阅增长最多的创作者                              |



## 高质量数据集：数据集实例

| 业务域    | 数据集        | 核心表                                                     | 核心企业知识                   | 经典用户场景                                        |
|--------|------------|---------------------------------------------------------|--------------------------|-----------------------------------------------|
| 广告域    | 广告资源位分析    | 广告位/资源位、<br>点击曝光转化、<br>业务线、<br>收入等信息                    | 广告相关术语/黑话等               | 开机大图广告位，近一周广告<br>收入趋势                         |
| 推搜域    | 推搜多维扫描     | 资源位<br>点击、曝光、转化/播放等表                                    | 资源位信息，<br>推搜相关内部知识库      |                                               |
| 客服域    | 客服数据集      | VOC、用户工单等                                               | VOC 内部术语等                |                                               |
| 技术平台域  | 数据治理       | OneMeta 表<br>数据治理表<br>数据增长表等                            | 表业务分域<br>负责人等            | 最近半年存储增长最快的top<br>10个表                        |
|        | Data+AI 平台 | 各数据产品运营表<br>数据任务配置、日志表<br>告警、监控信息表<br>ChatBI 全链路Trace表等 | 技术产品词汇、<br>成本组成<br>技术口径等 | 最近6个月AI用户理解大模型消<br>耗情况<br>每月ChatBI 访问UV、PV 情况 |
| 各垂直业务域 | ...        | ...                                                     | ...                      | ...                                           |

## 多智能体协同：Agent 架构



<https://lilianweng.github.io/posts/2023-06-23-agent/>

# 多智能体协同：各司其职，协同优化



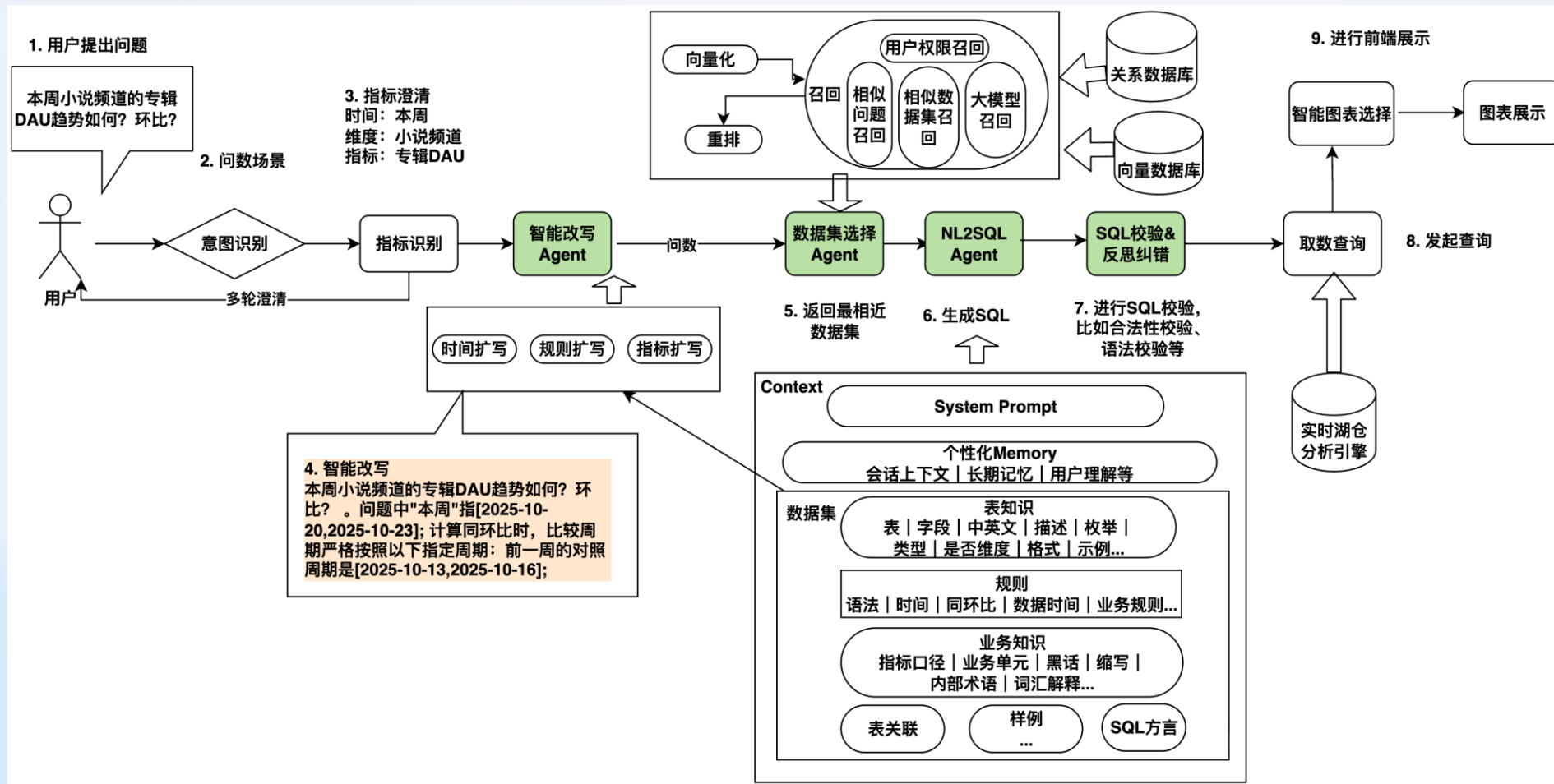
# ● 工具：Agent 手和脚

好的工具：安全、高效、便捷、可复用



# 多智能体协同：全流程拆解

一个例子：本周小说频道的专辑DAU趋势如何？环比？



# ChatBI 智能异动分析

## ChatBI MCP x Dify 构建人货场异动分析



# 落地效果

## 准确率

- 单数据集9个表，单表200+列，准确率 ~85%，人工复盘（非点赞/点踩）

## 效率提升

- 数据平权，数据分放和消费效率显著提升
- 人对数据的消费：取数效率百倍提升，从找数半天找不到→随时随地想问就问
- AI 对数据的消费：ChatBI 通过 MCP和API 提供给Agent 平台调用，支持自然语言接入，Agent 消费数据门槛大幅降低

ChatBI

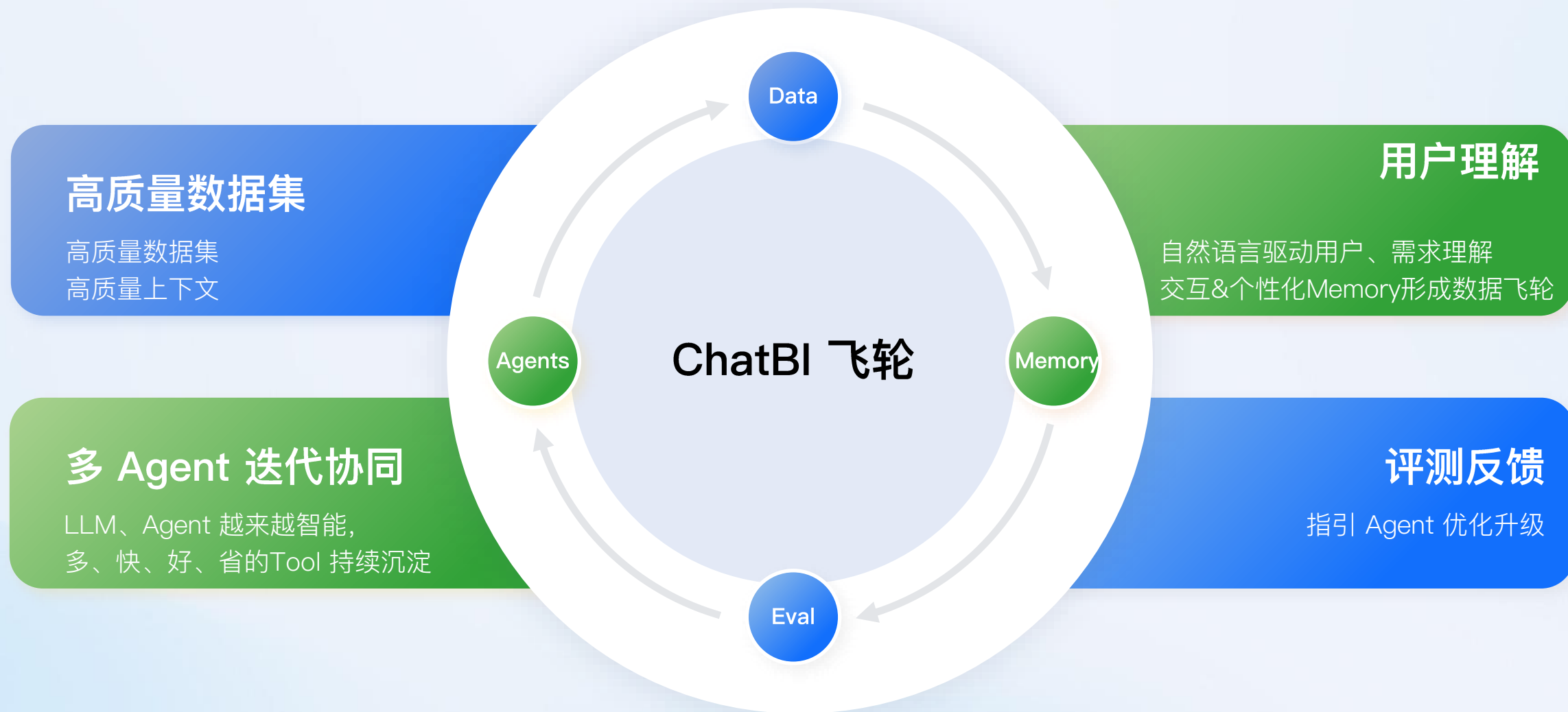
## 覆盖率

- 核心指标覆盖超85%，2万+ 智能问数
- 运营岗覆盖率~85%
- 产研、算法岗覆盖率~70%

## 数据价值释放、反哺数据建设

- 运营更爱用数了，聊着天就把活干了
- 自然语言驱动需求精准识别和及时响应，拉近与业务方距离
- 反哺数据治理：发现解决大量数据不规范问题

# 可持续进化的ChatBI



04

# 未来展望和规划



# 未来展望和规划

## 准确度提升

企业知识库持续治理和升级  
基于私有知识库和数据标注模型微调

## 人找数到数找人

支持订阅推送  
主动数据巡检、主动推送数据异动

## 数据深度洞察分析

Deep Data Research, 自规划深度数据洞察分析

📍 北京

**QCon**

全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

📍 北京

**AiCon**

全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

📍 上海

**QCon**

全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月

5月

6月

8月

10月

12月

📍 上海

**AiCon**

全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

📍 深圳

**AiCon**

全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

📍 北京

**AiCon**

全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOps
- 具身智能

# 极客邦科技 2025 年会议规划

促进软件开发及相关领域知识与创新的传播



参会咨询



查看会议

# THANKS

大模型正在重新定义软件

Large Language Model Is Redefining The Software

