

Multi-Agent架构驱动的 Data Agent路线与工程实 践

演讲人：李飞

数势科技 / AI负责人

AiCon

全球人工智能开发与应用大会

目录

01 Data Agent架构的演进路径

02 Data Agent 的核心能力

03 架构拆解与工程实现

04 实践经验分享

05 未来展望

📍 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

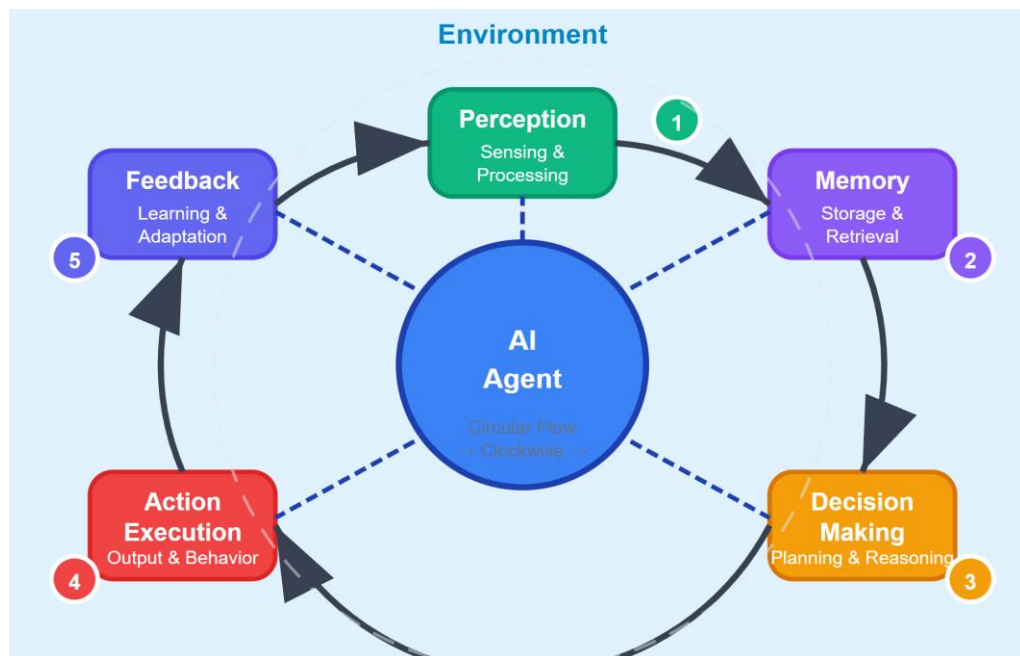
01

Data Agent架构的演进路径

单Agent到多Agent的进化

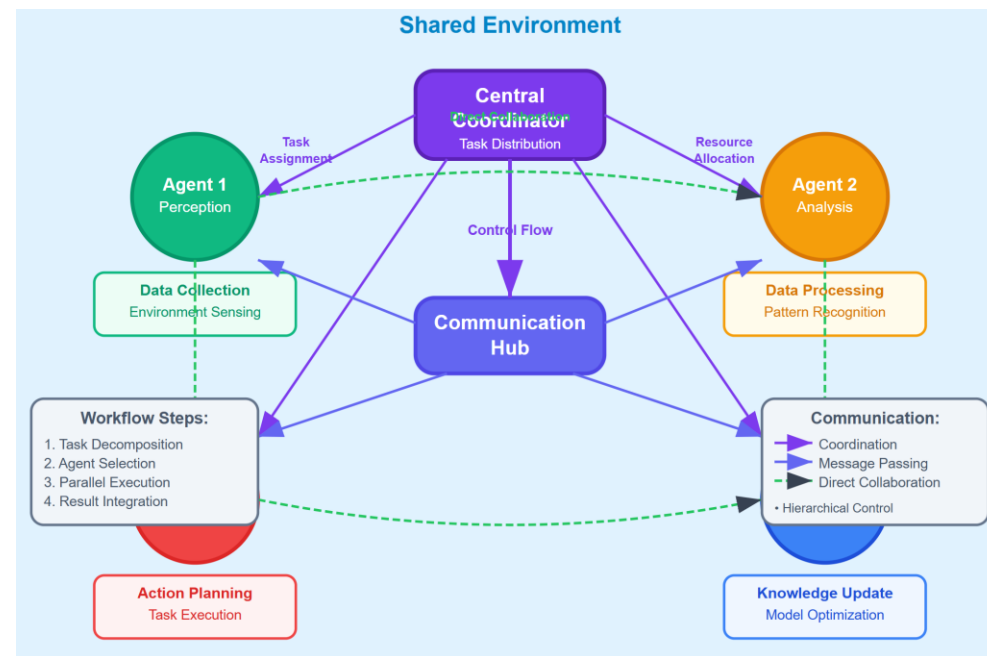
Agent通过引入规划、记忆、工具使用这三个核心功能，提高大模型落地的稳定性，从"好玩"向"好用"迈进。

单Agent



单Agent无法精通所有的技能，各个领域的规划执行流程不同，使用的工具也不尽相同。

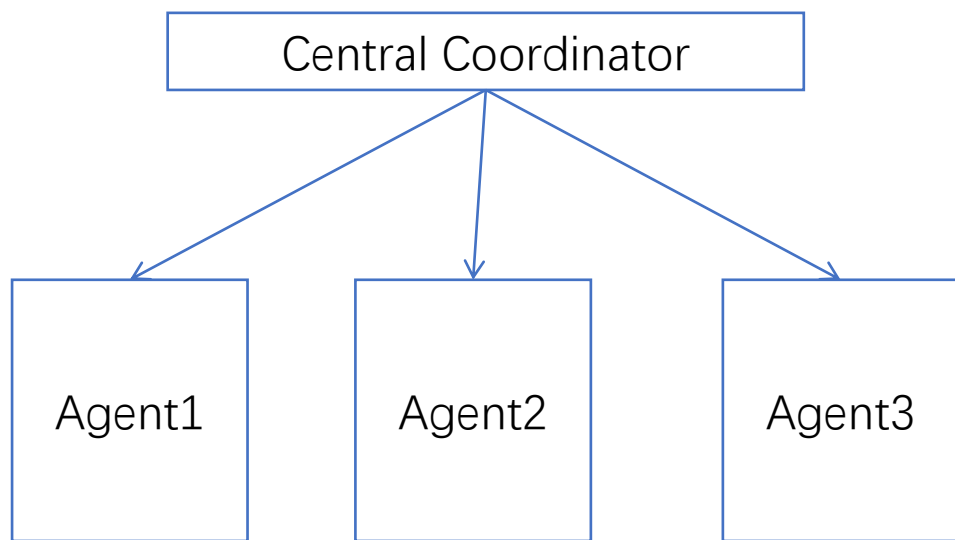
多Agent



多Agent降低工具和知识的搜索空间，保证任务聚焦和独立优化。

■ 中心化和去中心化智能体架构

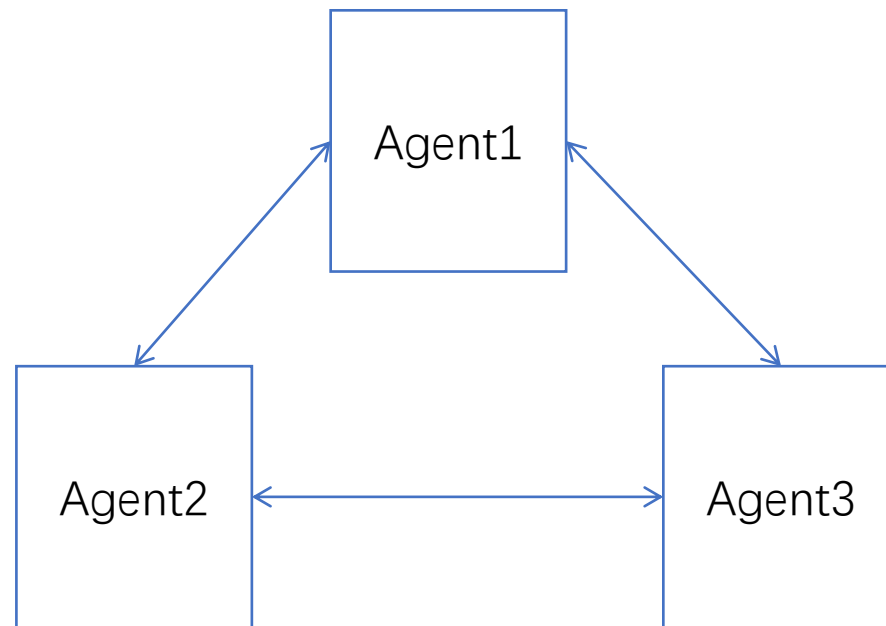
Coordinator统一接收所有Agent的信息，决策后再分发任务。所有Agent仅与中央节点通信。



AWS Bedrock MAC, ReAct等

强于全局优化，但面临单点瓶颈，例如coordinator

Agent独立决策，通过点对点通信协作，适合高实时性、隐私敏感场景。



斯坦福小镇, TAG等

扩展性优异（如斯坦福小镇的开放交互），但需解决非平稳性问题

Function Call vs MCP vs A2A

本质是快速“规模化”。

01

Function Call

Function Call不是协议，是大模型提供的一种能力，允许模型根据自然语言指令自动调用预定义函数，实现与外部工具的连接，但它并非统一标准——不同厂商对其接口格式、参数定义等有独立实现。

02

MCP

MCP是跨模型、跨工具的统一交互标准，不仅规范了工具调用（如函数描述、参数格式），还整合了 Prompts、资源管理等上下文体系。但生态成熟度不足，企业引入可能会造成运维负担，其标准化设计难以满足垂直领域的定制需求（如安全审计、数据隔离）

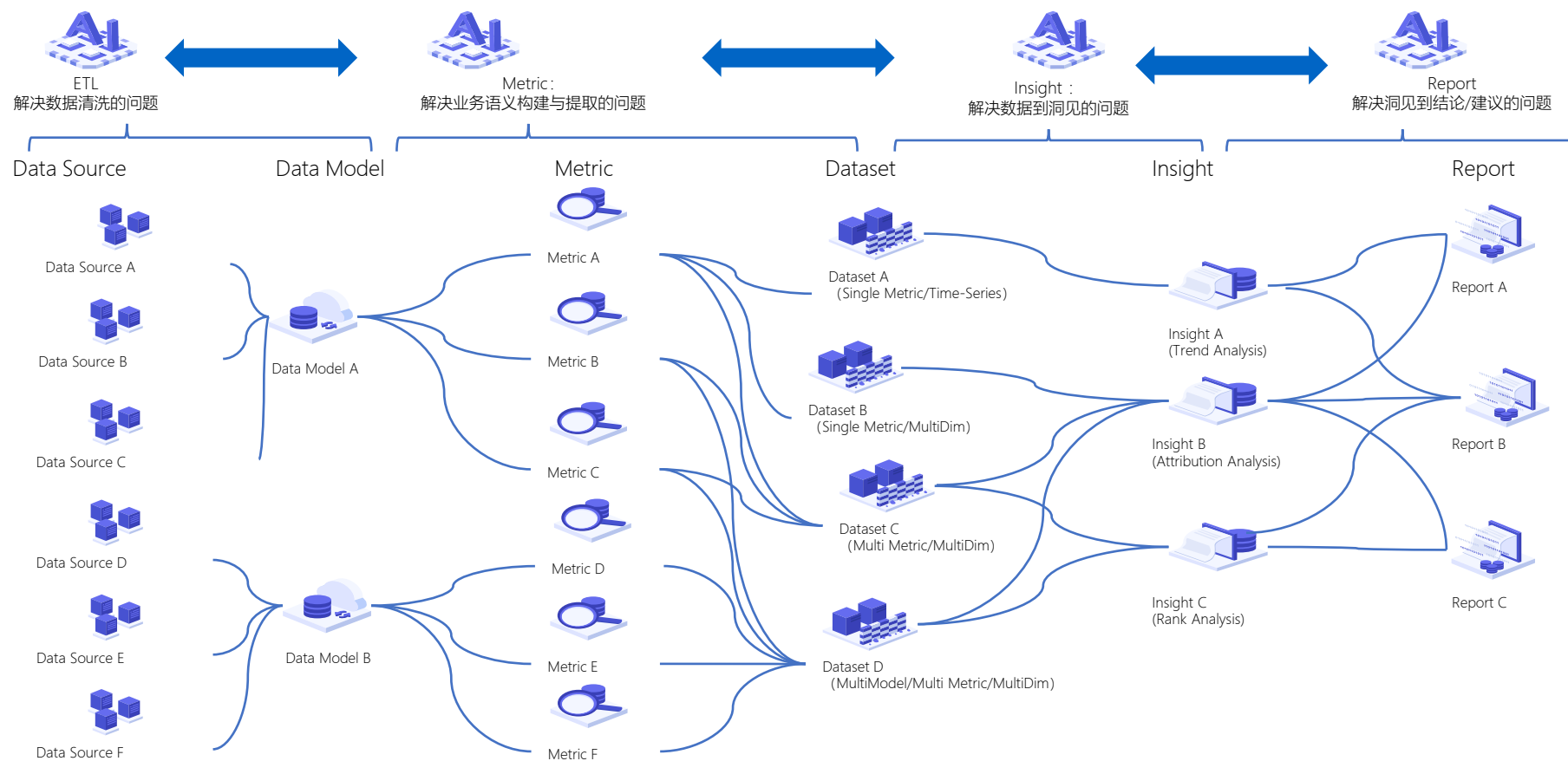
03

A2A

A2A则解决"多个Agent如何协作"，通过标准化Agent间通信，推动了多Agent系统的互操作和协作创新

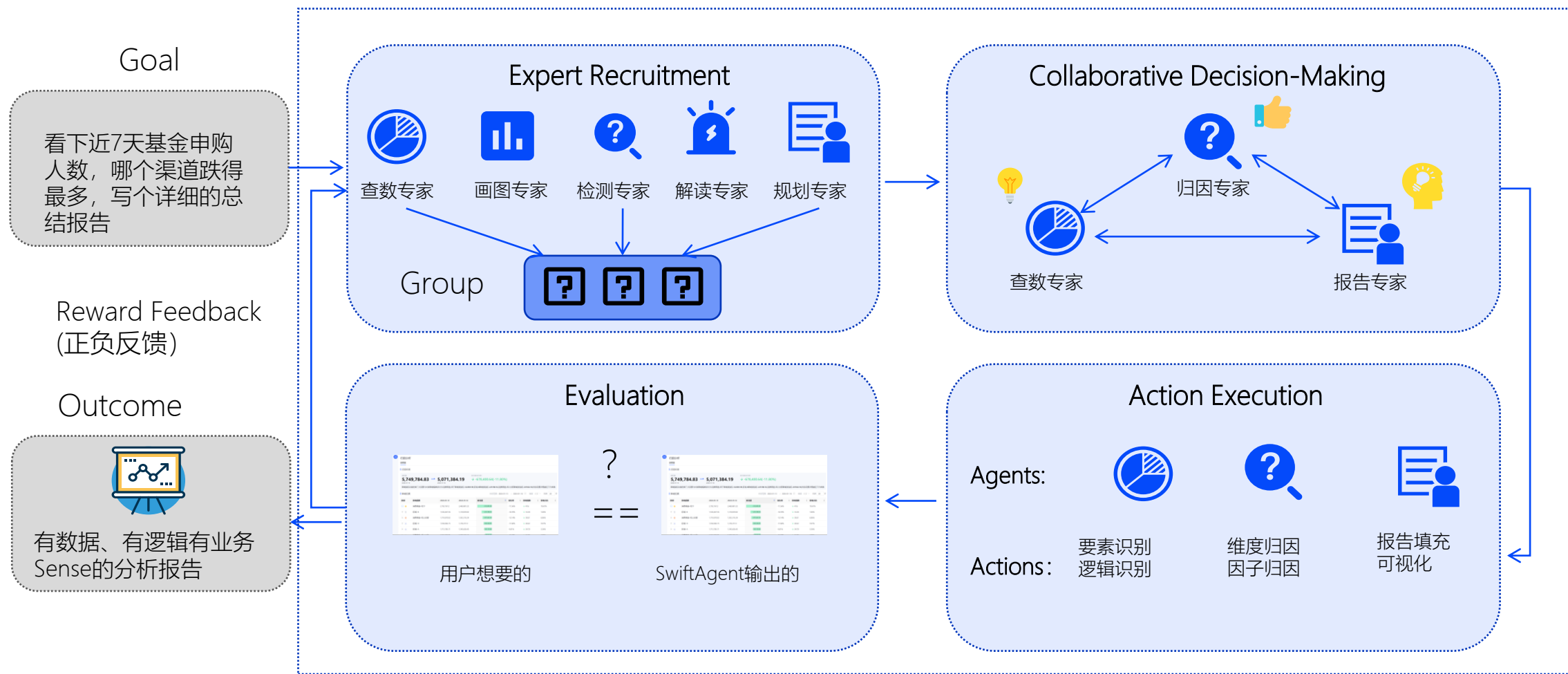
Data分析的全流程

“首先分析我们自己的收入指标增速，然后对比竞争对手的公开数据情况进行benchmark分析，然后结合websearch新闻，最后给我写一个完整的调研报告”



能像分析师专家一样规划： Data领域的“Multi-Agent”架构是如何运行的

N轮循环

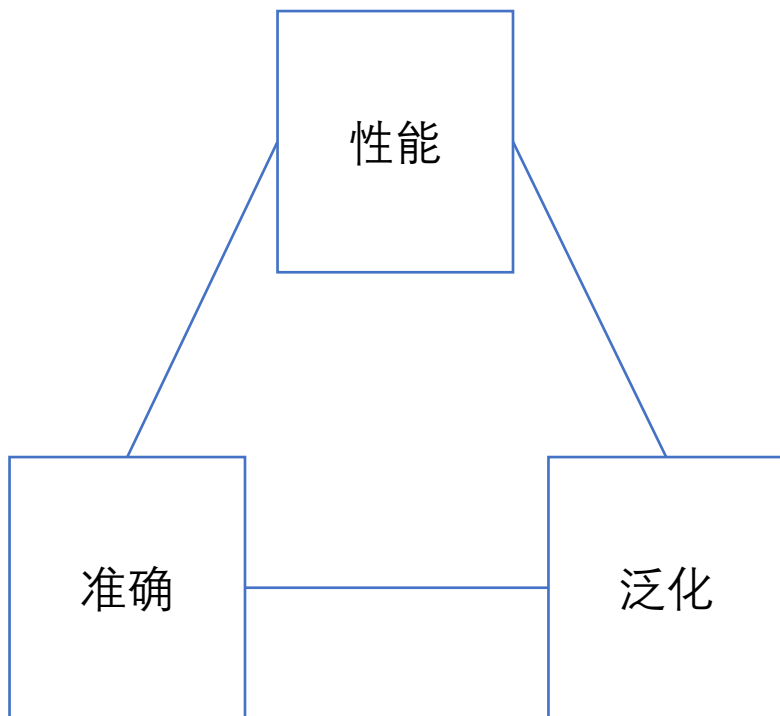


Data Multi-Agent的挑战



	优势	挑战
效率性	异步并行，可同时处理海量信息和任务，在复杂任务上效率极高	子agent之间无法直接通信与实时协作。
专注度	context压力相对较小，任务解决空间小，聚焦专注	在信息抽取和压缩过程中容易失真（例如数据解读是对dataframe结果解读还是通过算子编排计算解读）
扩展性	总信息量可扩展，无视上下文，领域能力扩展性强	子Agent的边界定义
复杂度	任务编排，异步任务管理、结果聚合逻辑设计复杂度高	依赖高质量的编排结果，消息状态管理和传递等

■ 当前多Agent的不可能三角



目标错位风险:

“100%正确率”和“绝对稳定性”本质是传统软件思维(确定性输出), 与大模型的概率性本质存在根本冲突。试图用蛮力对抗技术本质, 如同要求内燃机完全不排放废气。

边际效应陷阱:

从90%→95%准确率可能只需1个月优化, 但从98%-99%可能需要6个月, 而这1%的提升用户可能根本感知不到, 却消耗掉可支撑3个新功能的研发资源。

机会成本黑洞:

当团队所有精力被“稳定性攻坚”占用时, 其他产品正在定义新场景, 最终你的产品可能“赢了精度, 输了整个战场”。

——来自于DeepSeek

02

Data Agent的核心能力

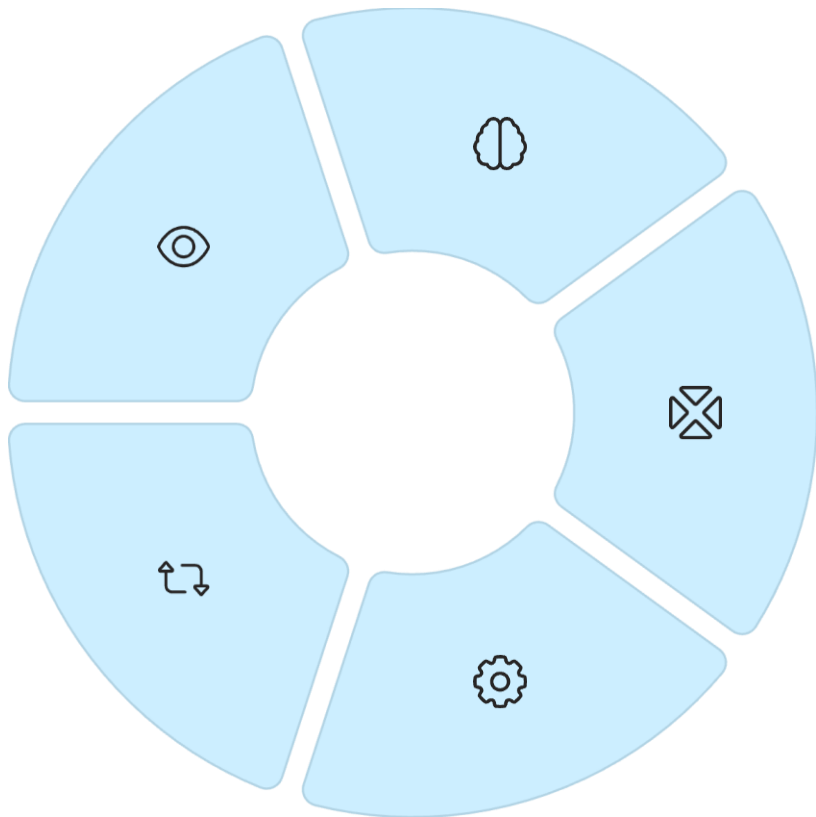
Data Agent的核心能力

多模态环境感知

Agent 能够感知并理解多种数据格式，包括结构化数据、非结构化文本、图像和语音等。它能够在复杂数据环境中建立上下文理解，识别关键数据实体和它们之间的关系。

持续学习进化

Agent 通过用户反馈和自我评估不断优化自身能力，记忆成功的分析模式并应用于新场景。它能够积累领域知识，提高分析效率和准确性。



动态复杂推理

基于多层次推理机制，Agent 能够处理因果关系分析、假设验证和异常检测。它通过逻辑推理和统计分析相结合的方式，从海量数据中提取有价值的洞察。

面向目标的行动规划

Agent 能够将用户意图转化为具体的分析计划，自动分解复杂任务，并根据执行结果动态调整策略。它能够平衡短期目标和长期价值，优化分析路径。

工具调用执行

通过抽象工具接口层，Agent 能够协调调用各种外部工具、数据库和模型。它能根据任务需求选择最合适的工具组合，并处理工具调用的异常情况。

Data Agent的核心架构



数据层：支持多源异构数据接入，提供灵活的数据集成方案，包括存算一体、存算分离和存算外置。

引擎层：数据语义模块通过精准的指标和标签平台，解决大模型的幻觉问题，并利用预计算加速引擎将复杂查询响应时间压缩至秒级，确保实时决策需求的满足。智能模型引擎模块结合大模型和小模型的动态协同。大模型负责语义理解与逻辑推演，小模型则专注于结构化分析。动态路由控制器根据任务自动分配。

应用层：为企业提供功能场景的重要支撑，通过工具组合，记忆共享，多Agent架构，实现智能体协同工作

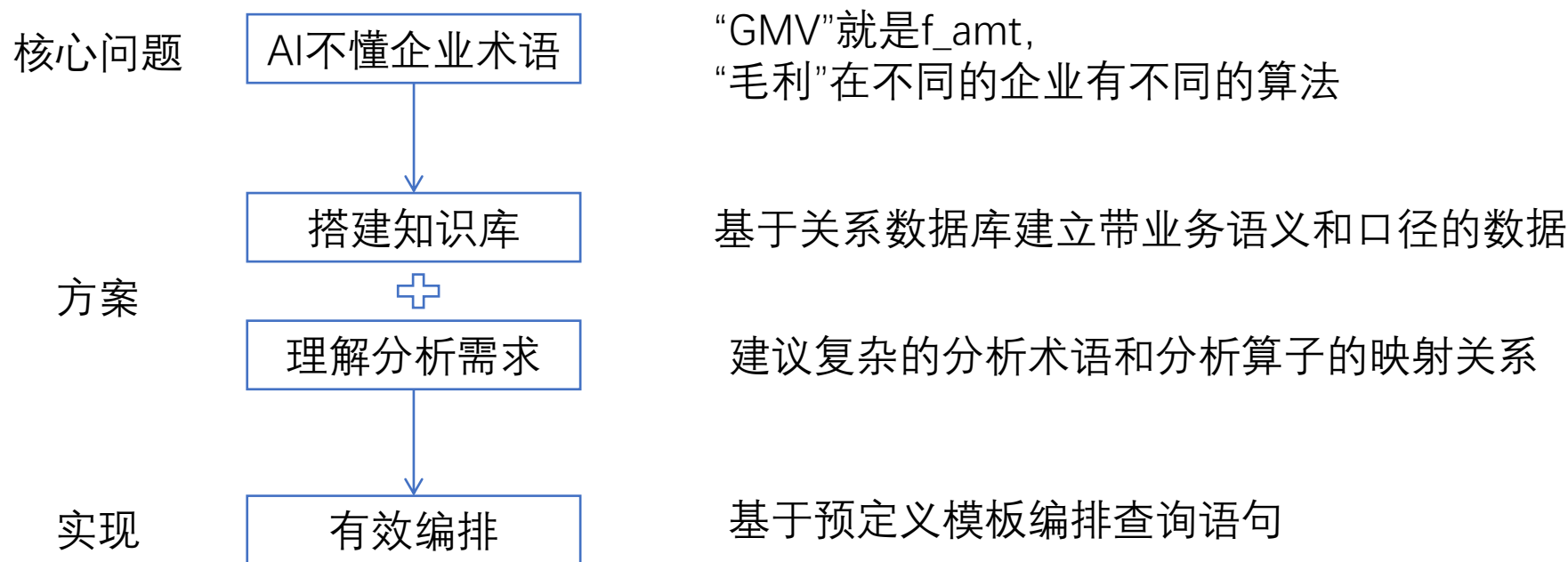
03

架构拆解与工程实现

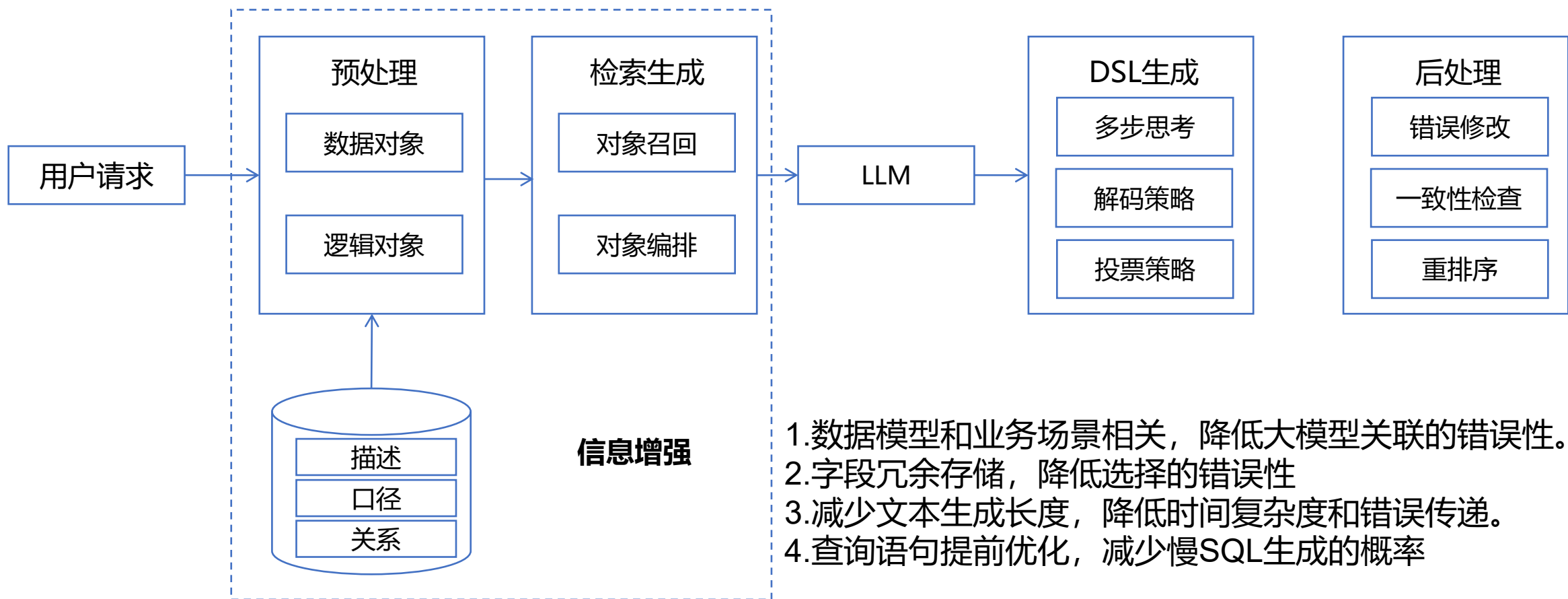


■ 企业级的数据查询和SQL生成的GAP

企业级的数据查询是让查数成为人和业务之间的“超级翻译官”



查数专家—NL2Semantics



- 1.数据模型和业务场景相关，降低大模型关联的错误性。
- 2.字段冗余存储，降低选择的错误性
- 3.减少文本生成长度，降低时间复杂度和错误传递。
- 4.查询语句提前优化，减少慢SQL生成的概率

业务动态适应的持续进化

指标市场

维度市场

问题模版

帮我看下昨天北京知春路门店的数据表现

发送



大模型：你到底要哪个！？

用户运营
部门理解

活动运营
部门理解

经营分析部
门理解

- “数据表现” = 新客首单人数、成熟用户复购人数、流失召回人数

- “数据表现” = 活动PV/UV、完成活动人数、活动带来的GMV

- “数据表现” = 门店营收、门店成本、门店利润、各类损耗指标

记忆即服务



参数化记忆

隐含存储于模型内部参数中的知识，通过预训练或后训练获得，嵌入在模型权重中，能快速检索事实性和常识性知识，但缺乏可解释性，难以选择性更新。

无需上下文，快速检索，提高推理效率。

但是

可解释性差
更新困难



上下文记忆

上下文记忆是显式的、外部的信息，用于补充模型内部参数知识，可进一步分为情景记忆和会话记忆，能支持系统动态存储短期和中期信息。

保障了模型输出的稳定性。

但是

存储和引用设计复杂

存储：动态更新，内容压缩，冷热分层，LRU缓存，
引用：重要度，时间衰减度，置信度

挑战：记忆一致性，性能，记忆敏感性

■ Memory-基本记忆操作

01

整合

将个体的短期经验经过加工与强化，转化为持久且可靠的长期记忆。

03

索引

通过高效的组织和分类机制，对记忆进行精细的索引管理，以便于后续的快速检索和利用。

05

检索

用户能够从庞大的记忆库中迅速访问到所需的相关信息，实现知识的灵活应用。

02

更新

系统能够根据新输入的信息，智能地修改和优化现有的记忆内容，确保记忆的时效性和准确性。

04

遗忘

智能地选择性抑制过时、冗余或不相关的信息，以保持记忆系统的精简和高效。

06

压缩

在保持记忆内容关键信息完整性的同时，通过保留显著信息和丢弃冗余信息来高效使用上下文

■ 记忆的更新和读取机制

【记忆更新】融合当前需求与全局目标

$$m_k^* = \arg \min_{m \in M} \alpha s^{rel}(m1, m2) + \beta s^{rec}(m1, m2) * \gamma s^{imp}(m1, m2) \\ + \beta s^{rec}(m1, m2) * \delta s^{rec}(m1, m2)$$

记忆更新考虑时效性，冲突性，扩展性(内容信息增强)和压缩性(内容信息减少)

记忆更新：冲突互斥，信息熵的增减。

【记忆读取】融合当前需求与全局目标

$$m^* = \arg \min_{m \in M} \underbrace{\alpha s^{rec}(q, m)} + \underbrace{\beta s^{rel}(q, m)} + \underbrace{\gamma s^{imp}(m)} + \delta s^{rel}(b, m)$$

综合考虑query和记忆的**最近性、相关性、重要性**，也要考虑用户背景和记忆的相关性

记忆权重：加权平均，时间平滑。

让Agent记住你是谁，你偏好什么。

Personalized Memory

个人字典 个人知识库 个人记忆

关于我 ▼ 介绍自己，帮助SwiftAgent给出更适合你的回答

开启 在聊天中应用

称呼 这将影响SwiftAgent与您对话时的称呼方式

岑润哲

所在行业 让SwiftAgent了解您需要的行业信息

银行

岗位 这将帮助SwiftAgent更好地理解您的业务背景

对公业务分析师

个人客户经理

对公业务拓展经理

柜面业务经理

信贷审批专员

组织 这将帮助SwiftAgent更好地理解您的业务背景

分析岗位

其他信息 可以是你的工作内容、关注的重点等

关于对公贷款余额增速

保存

Personalized Analytical Schema

查询员工总收入时，通常看最近一个月的。

编辑 删除

当我提到：查询产品类收入，一级品类名称。，我偏好问 最近一周。

编辑 删除

当我提到：营业收入，我偏好问 最近一周。

编辑 删除

当我提到：各个子公司的公募基金销售人数，我偏好问 昨天。

编辑 删除

Personalized Dictionary

添加个人字典

分类： 指标名称

目标值： 销售额 × 订单量 × 活跃用户数 × 日活 ×

描述值： 表现情况 ×

支持输入多个描述值,每次输入完成点击回车键完成创建

填写示例

原提问：

帮我看看昨天的 业绩情况

描述值 (指标的描述语)：

业绩情况

目标值 (指标的标准名称)：

销售额

订单量

实际查询：

帮我看看昨天的 销售额、订单量

记忆抽取的结构设计

结构化记忆

{ " 卖了多少 " : " 销售
额", " 上分 " : " 上海分
行", " 质量问题 " : " 发热,
易碎" }
.....



实体链接

借鉴GraphRAG，实现了记忆图谱用于增强表示记忆中实体与实体间的关系

非结构化记忆

会话偏好记忆：“用户查销售金额时候，喜欢使用饼状图进行对比。用户想要查的公司一般指上海银行。用户做比较的时候，也喜欢把北京分公司的对比加上”。

用户输入：2024年6月份，我的公司公募基金销售金额与5月相比下降？ ” }

.....



相关相似检索

记忆状态：增删改处理

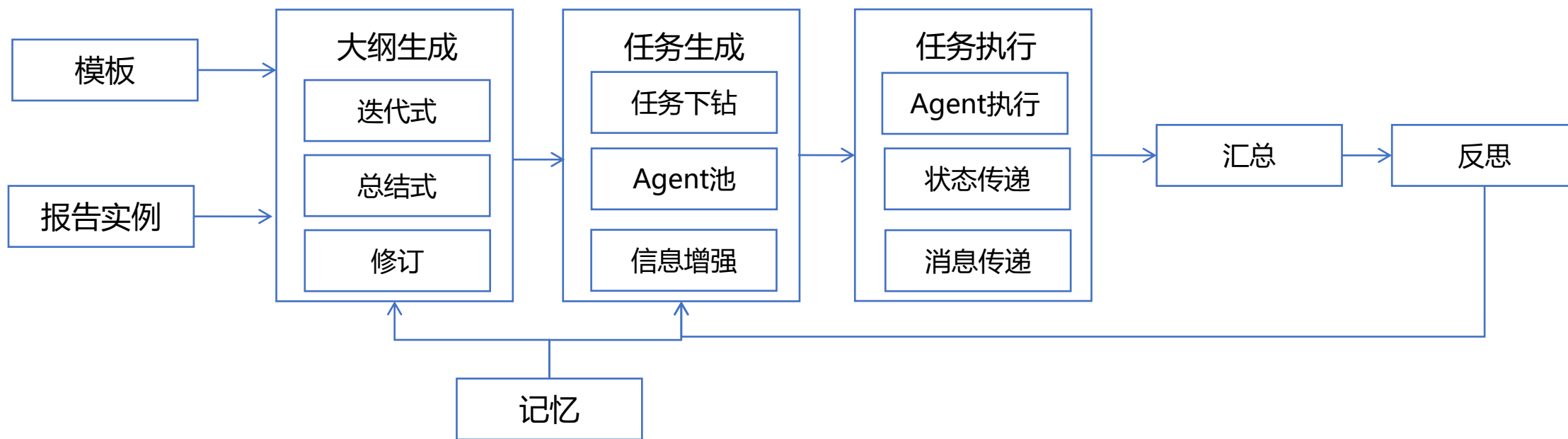
记忆权重：加权平均，时间平滑

- **时效性：**避免重复识别，提高计算效率；
- **准确性：**提高识别的效果和稳定性；

经常被使用的记忆会升级为长期记忆，而短期记忆会随着时间遗忘

Data Report实现链路

Research是Agent最好的落地场景，符合“慢思考”和“10”倍提效。

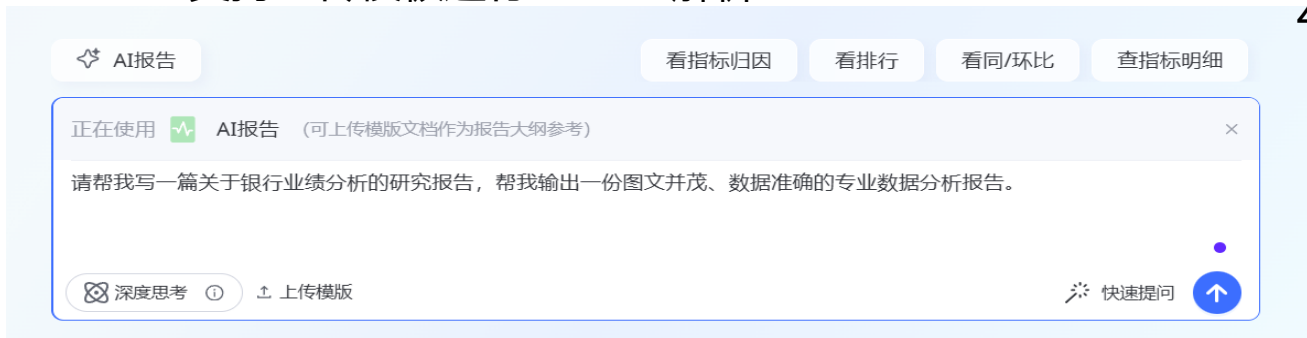


Data Report实现产品交互

3. 打印推理模型思考过程与大纲，大纲内容支持编



1.支持上传模板进行outline解析



2.用户也可以自由输入报告需求由大模型生成



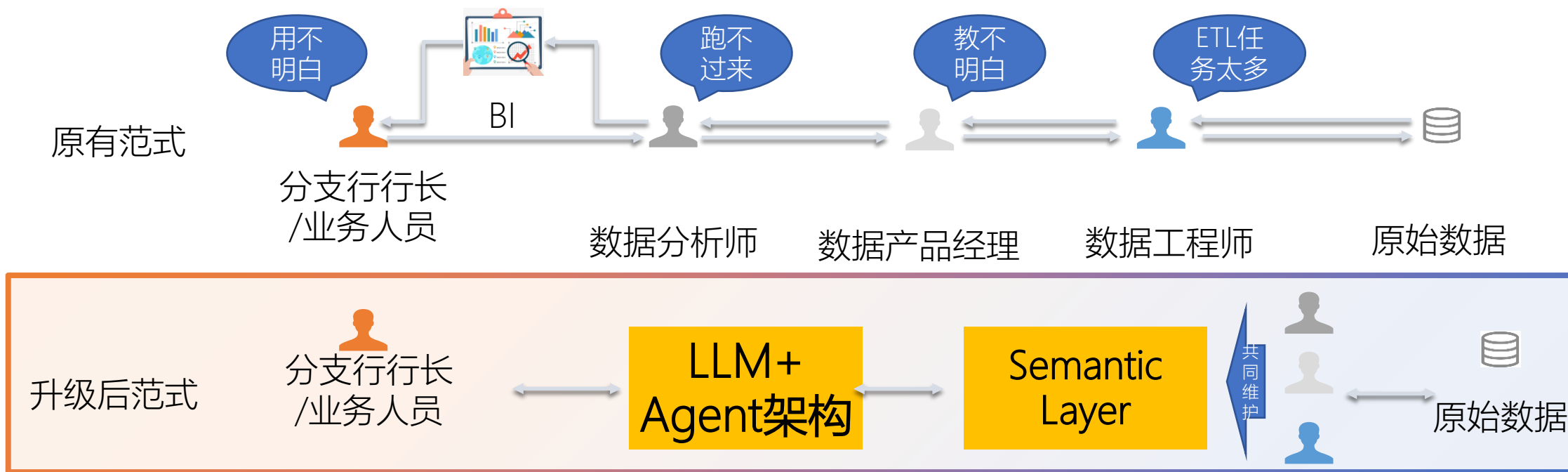
4. 生成图文并茂数据准确的分析报告



04

实践经验分享

■ 某头部城商行在数据分析层面的核心痛点



分支行行长/一线业务员直达数据，用数门槛降低，提升企业经营决策与日常业务流中数据参与度

■ 在某头部城商行支持的核心场景

以该金融机构的数据应用场景为例，通过**自然语言指标取数**，系统能够快速准确地获取相关数据；**智能归因分析**则帮助找出业务发展中的关键因素和问题所在；**自动报告生成**则以直观的方式呈现分析结果，为银行的决策制定、风险管理、业务优化等提供有力支持。这些场景展示了在银行业中运用自然语言处理和数据分析技术的重要性和实用性，有助于**提高银行的运营效率和决策科学性**。





“银行领导的一天”：如何利用AI洞察数据

关心的问题？



银行领导

1 指标异动推送

【对公贷款余额】出现指标异动，通知领导进行关注

2 多维度分析

我行今年每个月按分支行和客户类型的对公贷款余额如何？趋势是怎样？

3 归因分析

哪个分行下降的最多，哪些客户类型下降了

4 结论输出

帮我对上面数据形成一个总结报告吧

异动预警



归因分析



多维分析



数据解读



实现各类核心经营指标智能分析与报告生成

x50 ↑ 数据需求满足效率

90%+ 行领导分析需求满足度

100% 指标语义理解准确率

<6秒 从询问到结果

1200+ 上线一周领导主动询问数量

客户痛点



数据查询效率低：数据需求无限，开发资源有限，需求排队严重；



指标归因慢：主要依赖分析师人工在BI工具或Excel中进行维度与因子下钻，手工挖掘洞见



数据解读靠人工：领导每次对于大量表格表示看不过来，想直接知道结论和哪些指标出了问题

应用的核心产品能力



自然语言数据查询

行内领导可以通过自然语言形式**快速获取每日所需经营数据**，并支持智能生成可视化图表



归因分析

根据行内常用分析维度与指标的因子关系，进行自动化归因分析与建议生成，**帮助行领导快速发现数据异动的原因**，形成业务闭环。



报告生成

基于明细数据自动进行趋势分析、占比分析、排序分析和异常值分析等工作，并将生成的结果**以清晰易懂的报告形式**进行展现

使用场景



净收入分析



贷款情况分析



存款情况分析



同业负债分析

使用部门

项目一期服务重点



总行领导



分支行领导

项目二期服务重点



业务分析师

相关荣誉

知名数字化研究与服务机构沙丘社区发布《2025沙丘社区大模型先锋案例TOP50》，**数势科技凭借该项目入选此次先锋案例。**



■ 打造企业供应链分析助手，推动供应链团队订单管理效率提升

背景：

销售网络（线上线下多渠道并存，覆盖电商、大卖场KA、便利店等），供应链团队日常订单追踪和管理面临效率提升瓶颈，需要**高效洞察异常订单**，同时**针对订单满足率不达标KA客户进行归因分析**，从而挽回流失订单。

解决方案：使用 **SwfitAgent 智能分析助手产品**，打造企业供应链分析助手，推动供应链团队订单管理效率提升。

项目成果：

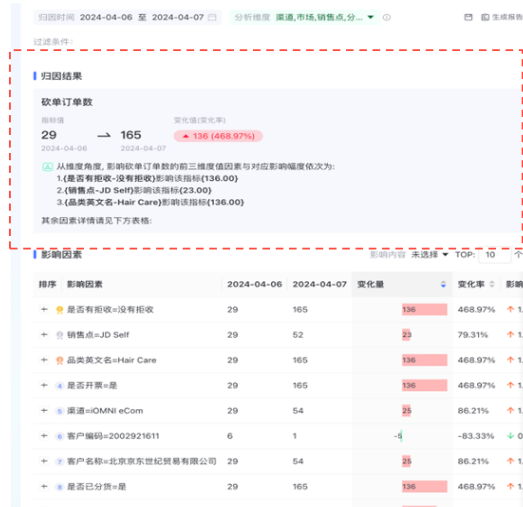
关键节点
供应链管理效率
提升150%

挽回可交付
订单商机
300万/月

1 语音问询SwiftAgent，迅速对比分析经营现状



2 识别异常订单，定位客户砍单原因



3 根因分析，并自动生成改善指引报告



作为店长数据秘书，实现全自动门店数据分析，提升运营效率

项目背景：某头部茶饮连锁客户，在推行数字化管理与决策，随着直营和加盟店数快速增长，门店督导管辖门店数量变多，人力成本极速上升

项目成果：第一周门店督导询问量突破**3000+**；分析工作处理时长减少**80%**；每人每周减少**10小时+**数据处理工作

1. 目标设定

明确巡检的主要目的，如提升服务质量、确保标准执行、优化门店环境等。确定要达成的具体指标和当前达成率。

● 目标设定

目标设定

● 取数与高级计算

达成情况排名

● 取数与高级计算

同店对照分析

4. 效果验收

对比巡检前后各项指标的变化情况。通过数据分析评估门店在服务、产品、运营等方面的提升效果。总结经验教训，为下一次巡检和持续改进提供参考。

● 取数与高级计算

门店前后对比

● 解读报告能力

总结报告

2. 巡店计划

制定巡店的具体地址和其它信息。确定每次巡店的重点关注指标以及波动原因。

门店信息检索

● 门店知识库查询

门店KPI查询

● 自然语言取数

门店KPI归因

● 指标维度归因

3. 门店稽核

检查门店的问题指标（如基础服务合格率、清洁卫生合格率、产品品质合格率、外卖点评合格率等），若有问题，从知识库中查找策略并推荐给加盟商

问题指标检索

● 自然语言取数

推荐策略查询

● 智库文档查询

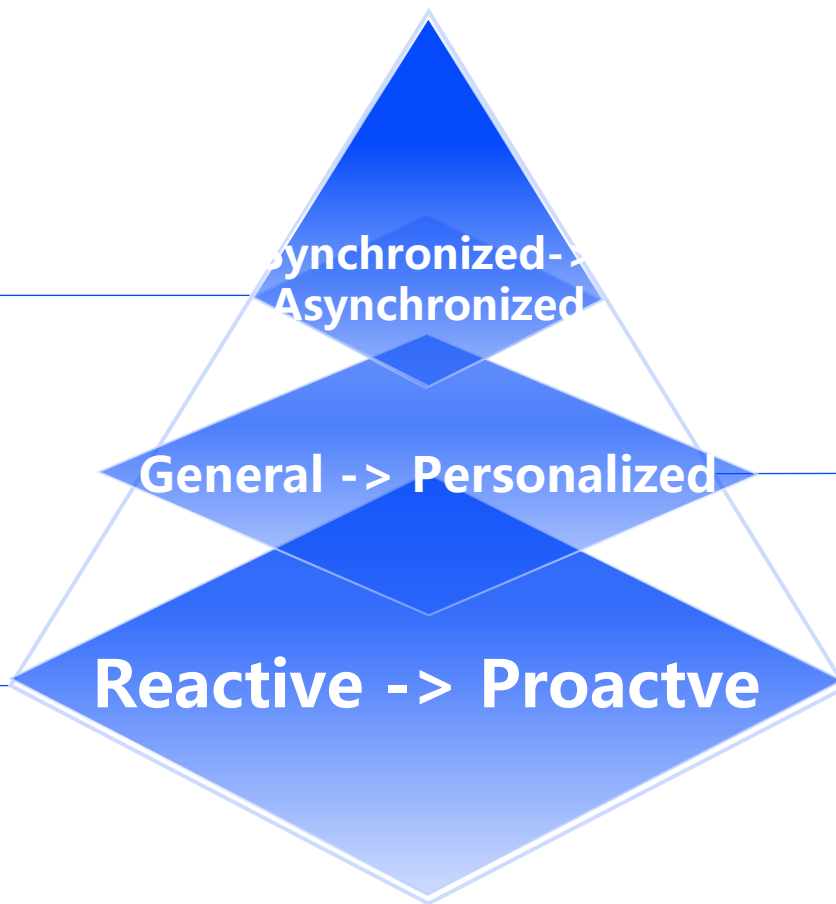
05

未来展望

■ 展望Data Agent的演进之路

Agent可以处理的任务将从简单同步型任务往异步复杂型任务演进

Agent将从回复用户问题升级为主动帮助用户发现问题



记忆模块将显著提升Data Agent对于用户个性化需求的理解与实现

📍 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

探索 AI 应用边界

Explore the limits of AI applications

AiCon

全球人工智能开发与应用大会

感谢您的观看

公司和个人介绍

数势科技为企业提供基于大模型增强的**数据分析 AI Agent**，提升企业的数字化决策能力，推动企业数字化升级。

公司概况

- 2020年，由原京东技术副总裁黎科峰博士创立
- 行业首个商业化落地的**数据分析智能体SwiftAgent**
- 在金融、零售、先进制造领域有深度技术和业务know-how

部分代表客户



投资机构

