



# 金融大模型数据治理与应用创新

2024年8月

密级：

# 目录

## CONTENTS

1

金融数据的生产现状分析

2

AI原生下的金融数据生产

3

数据治理领域的AI技术应用

4

数据中心下的大模型消费应用探索

# 1. 金融数据相关的指引和规范

## 监管文件：

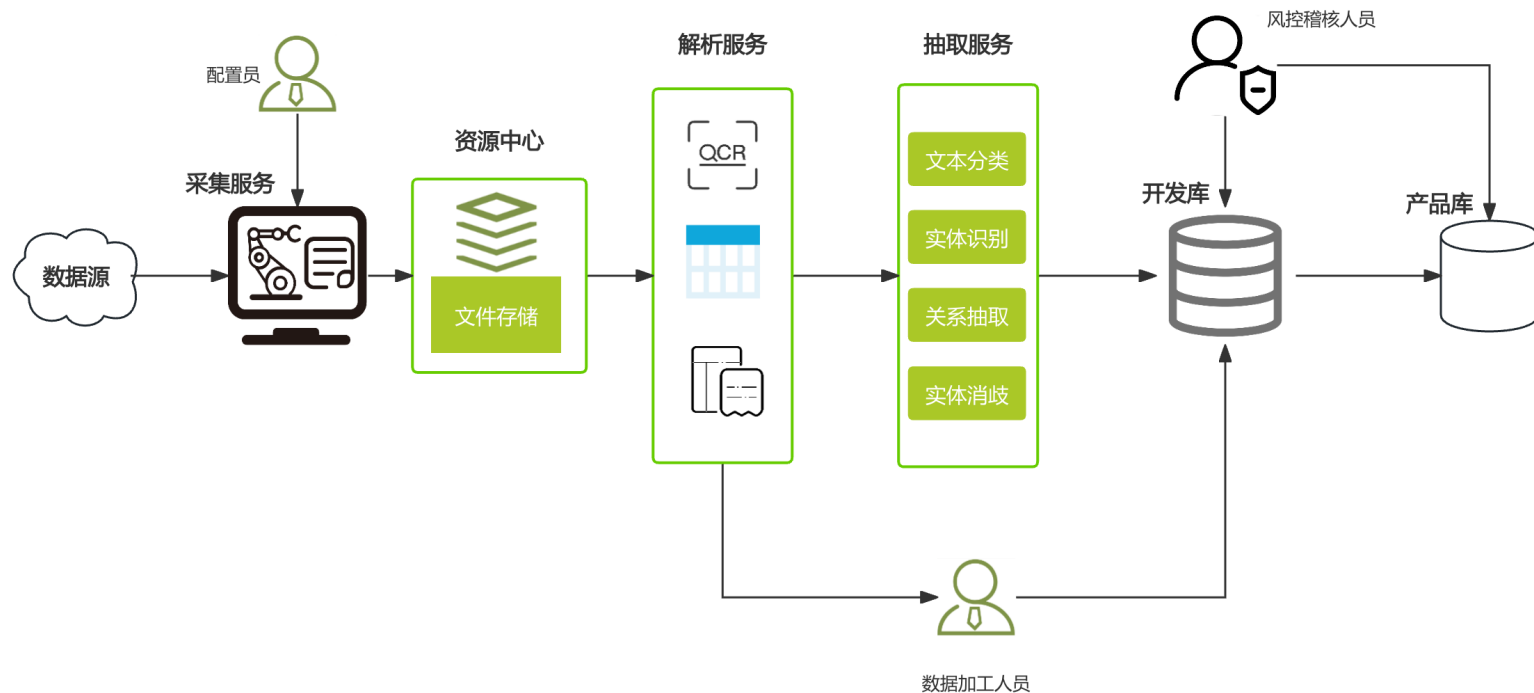
- 《人工智能风险管理指引》（中国银保监会，2022年）：该指引针对金融机构在应用人工智能技术时的风险管理进行规范，旨在确保人工智能应用的安全性和合规性。它提出的原则和要求包括：数据隐私保护、算法透明度、模型验证等方面。
- 《金融科技创新监管试点指引》（中国人民银行，2023年）：该指引旨在推动金融科技创新，并在监管方面提供指导。它明确了金融科技创新监管试点的范围和原则，鼓励金融机构与科技企业合作，推动金融科技创新发展。
- 《国家金融监督管理总局关于银行业保险业做好金融“五篇大文章”的指导意见》（国家金融监督管理总局，2024年）：该意见明确把科技金融、绿色金融、普惠金融、养老金融、数据金融作为发展新质生产力的五个关键领域。

## 行业趋势：

- 鼓励金融数据的生产加工和消费环节引入AI技术，努力寻求促进创新和控制风险的平衡点。
- 数据在金融科技创新中的关键作用在AI技术爆发背景下越来越处于中心地位。
- 数据金融本身开始从幕后走向前台，开始展现出独立价值和影响。

密级：

## 2.金融数据生产的关键要素分析



AI辅助下的金融数据生产流程示意图

密级：

## 3.金融数据生产问题分析

### 1.生产成本:

- 单人工成本逐年升高，对时效和数据质量的要求进一步要求更多的人力投入，技术进步会让单位算力、存储、网络等成本持续下降。

### 2.数据时效:

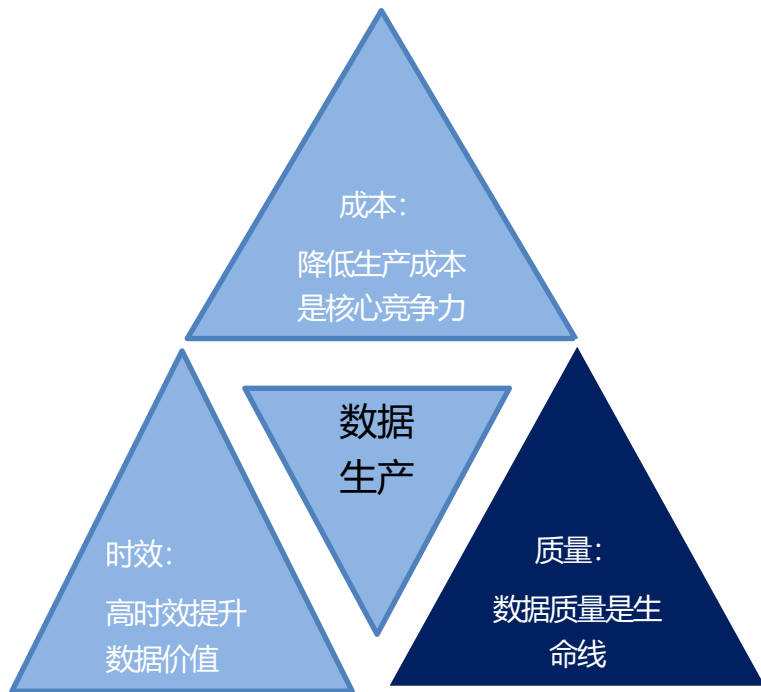
- 数据时效是妥协的结果，成本可控下始终有提升数据时效的必要性。通过提升自动化率、优化人工排班、分级处理等提升时效的空间已经不大。

### 3.数据质量:

- 通过人工审核、人工双录、稽核校验、问题专项等方法保障准确率目前已经到了一定瓶颈，自动化工具引入不能覆盖所有数据类型。

密级:

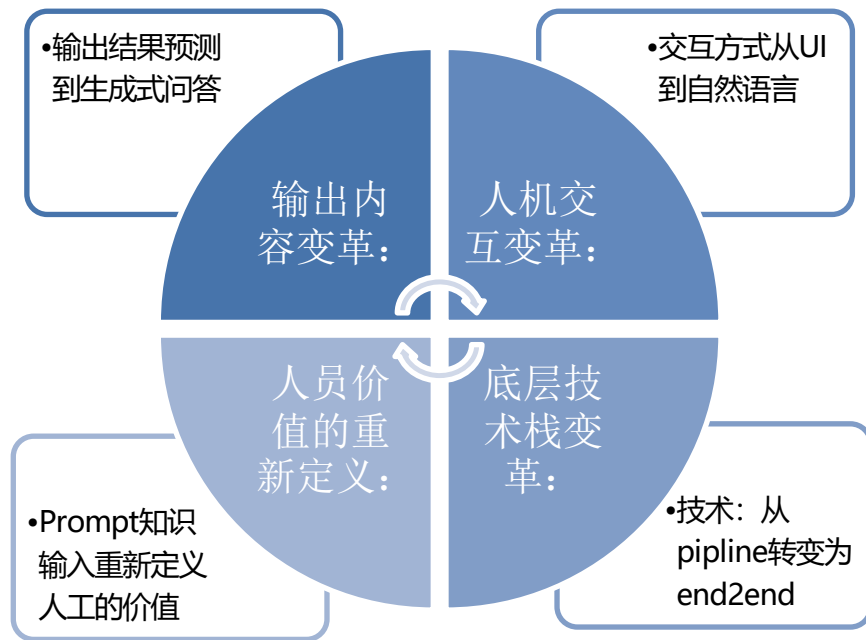
以人工为中心构建的金融数据生产模式，引入AI辅助提升自动化水平已经到瓶颈。



## 4.AI辅助到AI原生下的数据生产方式转变

AI原生是以AI为中心来构造数据管理、业务流程、人机交互、产品设计。

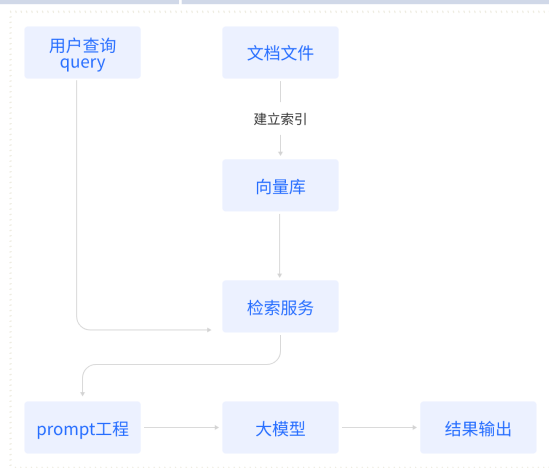
- AI在应用中类似大脑承担中心决策的角色，AI原生应用表现为类人智能体，衍生特征则是智能优先。
- 数据产品的设计、工作流的输入输出接口、评价机制和数据消费都需要从以人为中心转变为以AI为中心。



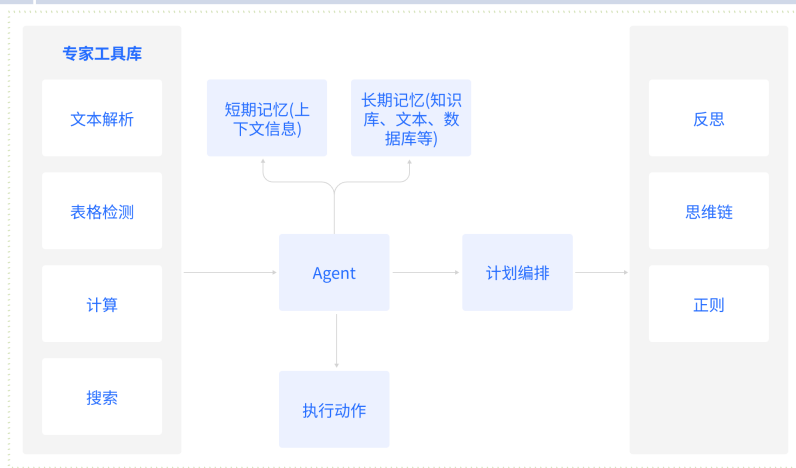
密级:

## 5.大模型应用下的RAG和Agent

对比项	RAG	Agent
应用阶段	微调和推理	当前仅推理
价值体现	放大大模型的能力	丰富大模型的能力
适应场景	垂直领域，有限空间下突破准确率临界点	覆盖多种模态和功能点，实现端到端AI能力

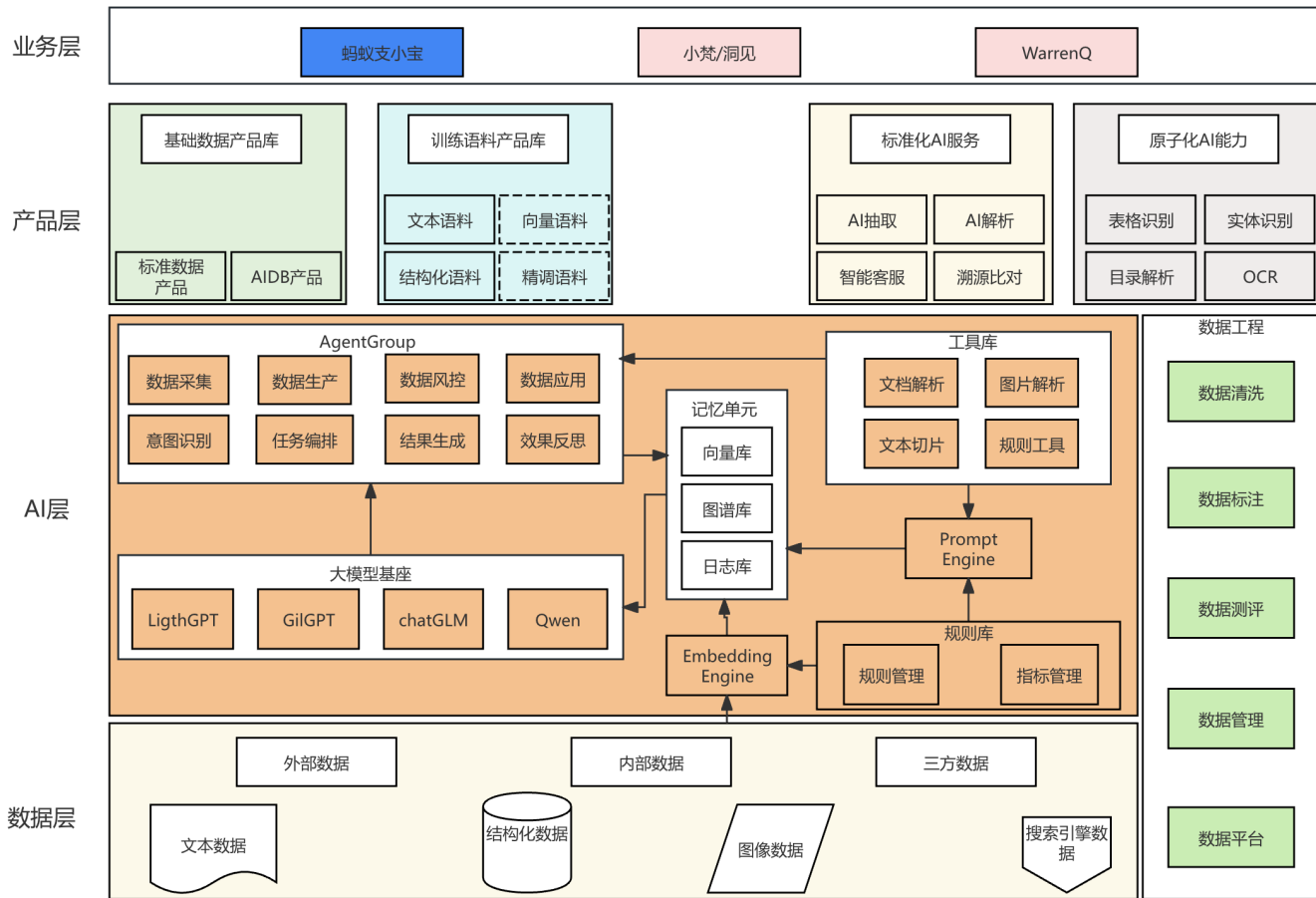


RAG原理与结构示意图



Agent原理与结构示意图

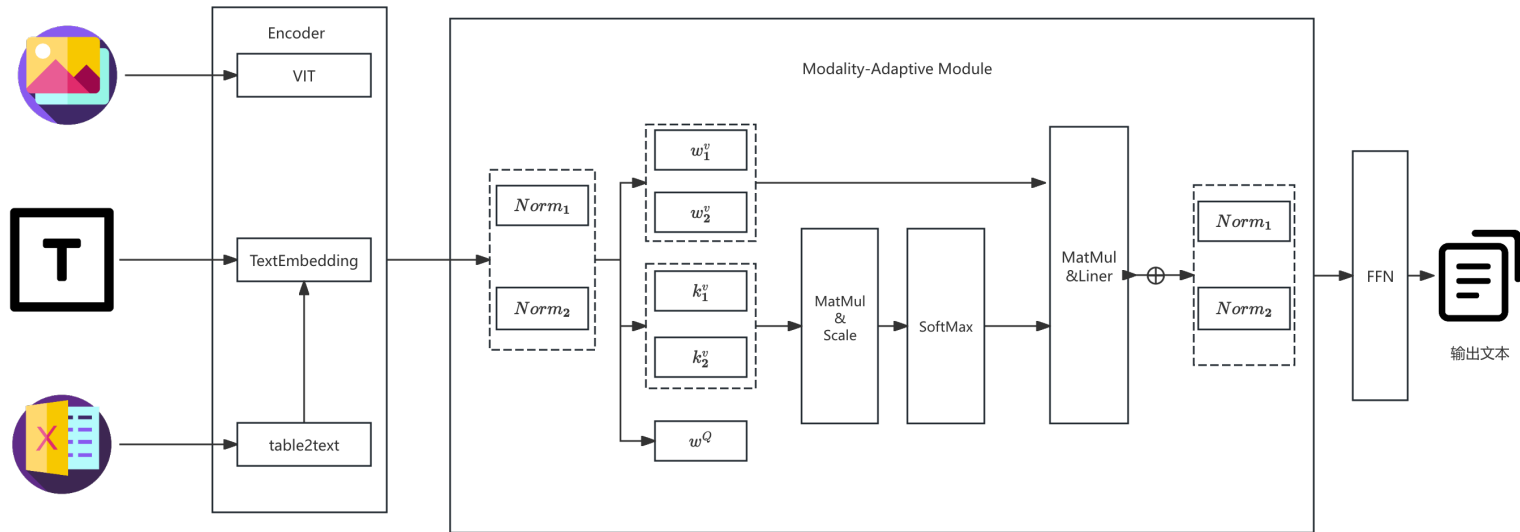
## 6. AI原生下的金融数据生产架构



密级：

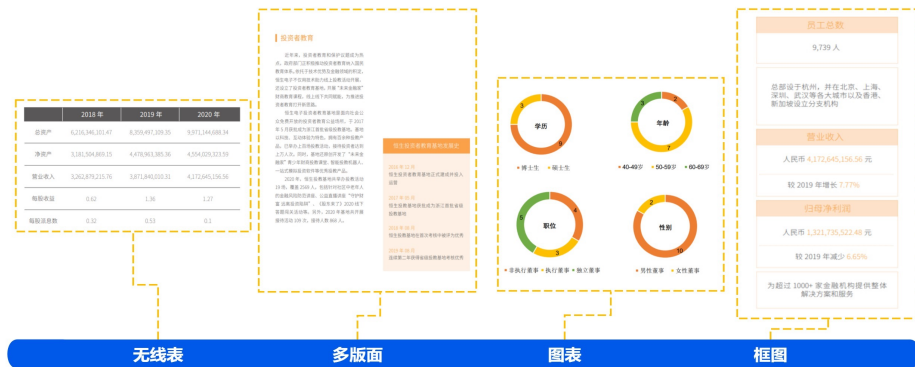


## 7.多模态文档理解

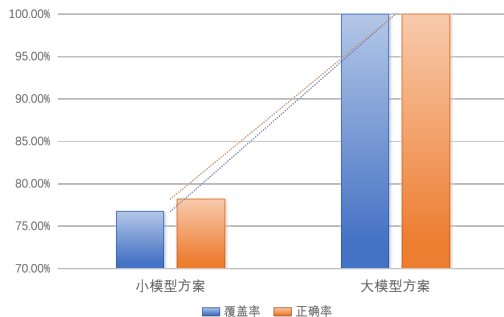


多模态图文&表格理解模型架构

## 8.多模态文档理解

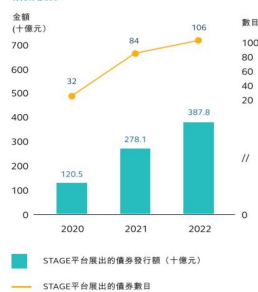


ESG报告中的多模态文档形式展示

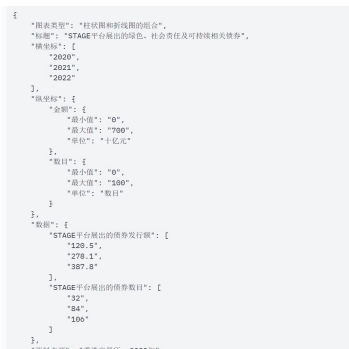


ESG报告的多模态理解效果

STAGE平台展出的绿色、社会责任及可持续  
相关债券



資料來源：香港交易所，2022年



ESG报告中的图表理解

密级：

## 9.数据治理领域的AI技术应用

数据是土地、劳动力、资本、技术之后随着信息化时代兴起尤其是在AI时代后公认的第五大生产要素。  
数据和AI大模型是一个非常好的相互促进的闭环场景。

数据治理面临哪些问题？

### 1.只看局部不看整体

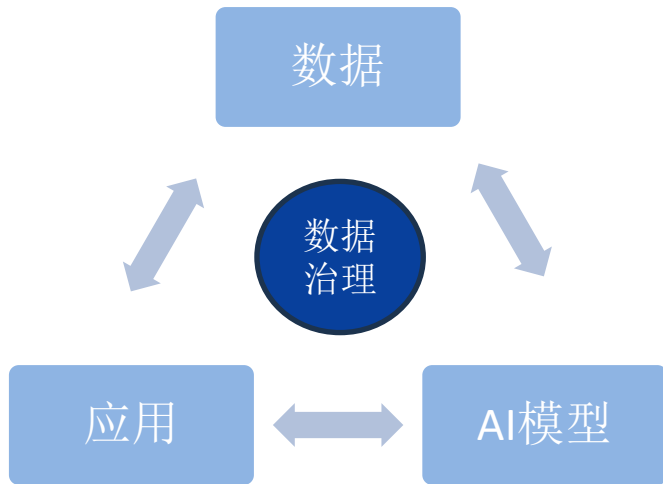
- 视角狭隘，认为数据治理就是“数据清洗”，对数据治理缺乏全局视野和整体方案设计。

### 2.不重视“最后一公里问题”

- 数据治理本身不是目的，治理后怎么用以及产生足够的收益是关键。

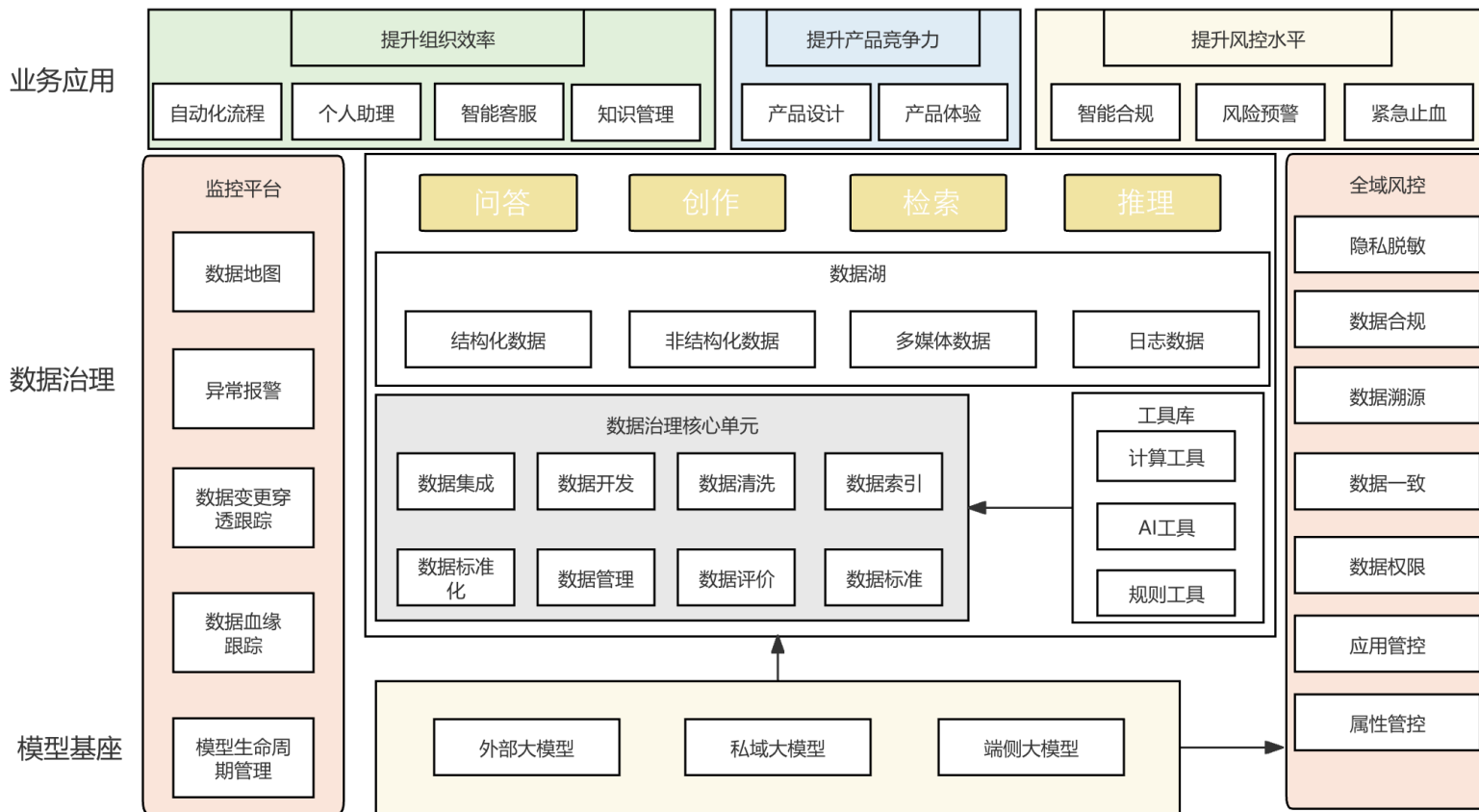
### 3.数据安全性考虑不足

- 缺乏对数据的流转、存储、检索、清洗、应用各环节的可信隐私保护和合规设计。



数据治理和AI模型的关系

# 10.AI大模型下的数据治理架构图



密级：

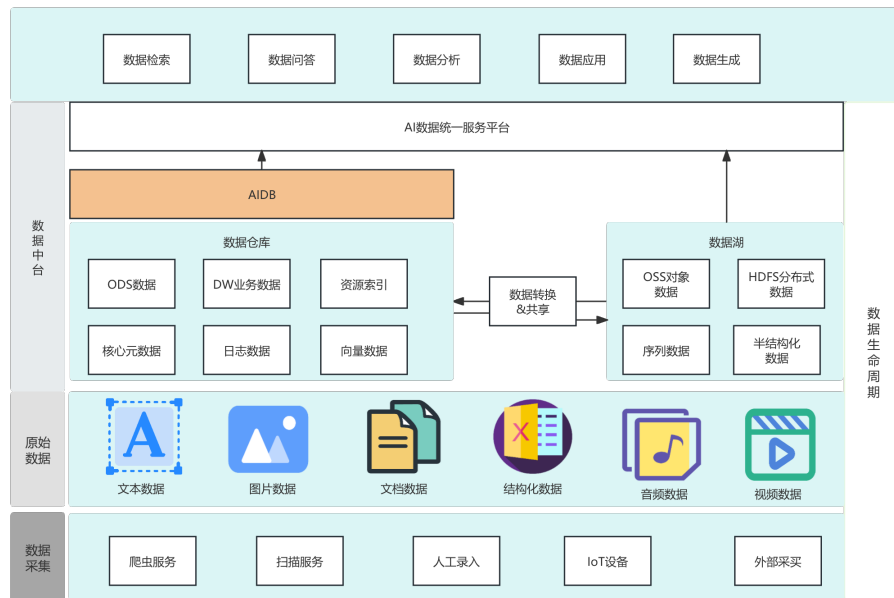
# 11. “湖仓一体” 的AI友好型数据存储模式

“湖仓一体” (Lakehouse) 是一种新兴的数据存储架构，它结合了数据湖 (Data Lake) 和数据仓库 (Data Warehouse) 的特点，旨在提供一个统一的平台来存储、处理和分析大量结构化和非结构化数据。

## 要解决什么问题？

- 1.数据湖提供了灵活性和扩展性，但缺乏数据治理和优化的查询性能。
- 2.数据仓库提供了结构化和优化的查询性能，但缺乏处理非结构化数据和快速迭代的能力。
- 3.面向AI的原生数据架构设计需要满足多种类型数据的统一服务。

密级：



## 12.企业级私域大模型数据治理应用

私域大模型的  
兴起得以让数  
据治理的价值  
最大化。

数据  
工程

拟人类AI

生产力类AI

风控类AI

通用企业级私域大模型：

- 1.数据整合与统一数据视角：
- 2.数据隐私和安全性
- 3.模型性能与可扩展性

金融类企业级私域大模型：

- 1.严格的合规性要求
- 2.模型的可解释和透明化
- 3.高度的准确性和一致性

数据产品化

Agent产品化

模型训练微调

专家助手建设

大模型基座

RAG构造

Agent框架

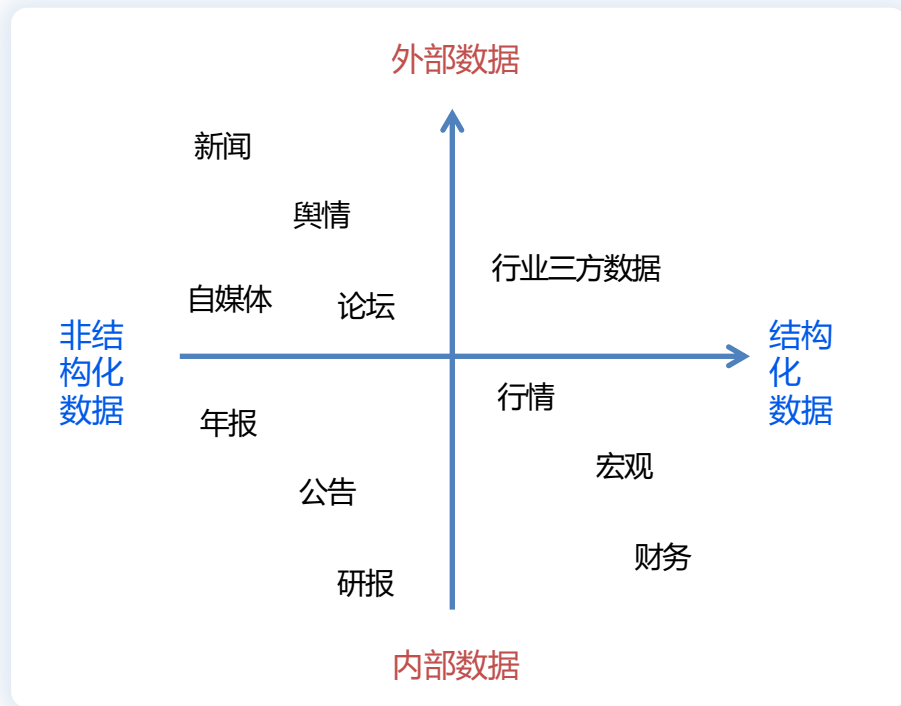
密级：

# 13.以数据为中心构造AI应用

以数据为中心（Data-Centric）和以模型为中心（Model-Centric）的大模型设计范式是两种不同的人工智能系统构建方法，它们在设计理念、工作流程和优化重点上有所区别：

比较项	以数据为中心 (Data-Centric)	以模型为中心 (Model-Centric)
设计理念	强调数据的重要性，认为数据质量和利用效率是模型有效性的关键	认为模型架构和学习能力是系统性能的决定性因素
工作流程	从数据采集开始，经过数据预处理，然后是模型训练和评估	首先集中在模型选择和设计上，然后是数据适配和调整
研发重点	从业务、格式多样性、数据分类、打标、清洗、标注等维度形成完整数据工程处理方案	聚焦模型架构，包括网络深度、宽度、激活函数和正则化技术、微调方法、embedding方法等
数据依赖性	对数据依赖性高，更多数据通常意味着更好的模型性能	模型复杂性可以部分补偿数据不足，强大的基座模型获取能力外溢
维护成本	易于扩展，随着数据工程的精细化，数据管理成本会增加，模型研发成本会降低	数据管理无需投入过大精力，模型更新和维护可能更复杂，成本增高

## 14.数据工程分析方法举例



### 数据工程能力

#### 1. 数据表达能力：

通过数据切片、向量化、ES召回、实体抽取技术，形成输入大模型前的数据表达服务能力。

#### 2. 数据服务能力：

建立NL2API、NLP2SQL的数据服务，形成大模型对数据的消费能力。

#### 3. prompt 工程能力：

通过业务建模，设计提示模版，建立AI模版测评能力。

密级：



# 15.Multi-Agent下的智能投顾

实现从AI原子能力到AI系统能力突破，以领域专家经验利用私域和公域数据获得高价值答案。

## 核心观点

亚世光电近期的表现可以从其估值水平、亚世光电的

## 新闻资讯

【利好消息面】从于光学光电子板块的强的火热发展。作为虚拟有广泛的应用，随着虚拟展潜力巨大。此外，亚势，吸引了投资者的关注好消息的细节需要进行一

## 公司亮点

市场对其电子得知具体的和的基本面情况光电在VR/AR该技术的期待

## 基本面分析

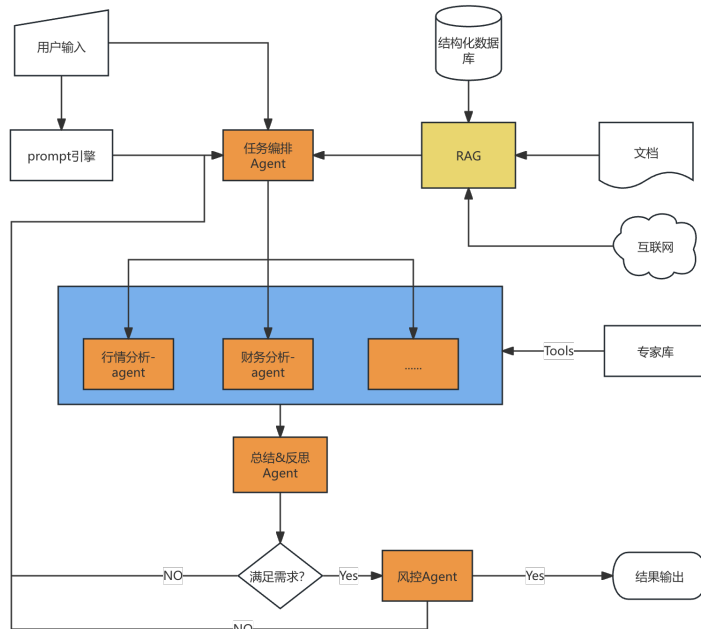
根据聚源数据，亚世光电在基本面速览维度下，有几个关键信息：1. 最新交易日PE在行业排名前10%，PE估值处于较高水平。这意味着该股票相对于同行业其他公司来说，其PE值较高，投资者可能需要谨慎考虑其估值情况。2. 最新交易日PE和PB均处于历史70%分位以上，即处于较高区间，这表示该股票的历史估值大部分时间都处于较高水平。3. 最新交易日预测PEG小于等于1，这意味着市场对该股票的估值可能低于其成长性，未来可能会有较好的增长前景。4. 根据最新财务报告，归母净利润同比下降大于等于30%，表明公司归母净利润出现了大幅下降。这可能会对公司的盈利能力产生一定影响，需要关注其业绩变化的原因。这些关键信息综合表明，亚世光电在基本面速览维度下，其

## 个股排雷

根据聚源数据，亚世光电在光学光电子行业中的关键信息包括其资金净流占比和资金净流入。具体表现为，在最近一次统计中，亚世光电的资金净流占比和资金净流入数据尚未提供。这些数据可以反映出该行业对亚世光电的资金态度和偏好。从行业排名的角度看，虽然无法直接通过数据得知亚世光电在光学光电子行业的具体排名，但结合其资金净流占比和资金净流入情况，可以推测其在该行业可能具有一定的竞争力。这些关键信息表明，亚世光电可能吸引了部分投资者的关注，并在光学光电子行业中占据一定的市场份额。【4】

密级： 个股排雷

明近期亚世光电的交易情况较为稳定，市场对其关注度较高。综上所述，亚



Multi-agent框架下的技术架构图

HUND SUN

让金融变简单

THANKS



恒生电子股份有限公司  
HUNDSUN TECHNOLOGIES INC.  
+86-571-28828888  
[www.hundsun.com](http://www.hundsun.com)

密级：

版权（报告、PPT、演讲内容）仅为恒生电子所有，未经书面许可，任何机构和个人不得以任何形式转发、翻版、复制、刊登、发表或引用。