



基于改进型RAG技术的金融研报智能体

队伍名字不能为空 · 2025年8月

Contents

目录 ▶

01

团队介绍

02

设计思路

03

解决方案

04

方案成效

05

技术创新

06

待改进点

01 团队介绍

队伍名字不能为空

赛题四：智能体赋能的金融多模态报告自动化生成



02 设计思路

「 从Multi-Agent到改进型RAG 」



01 目标

构建自动撰写 **宏观研报** **行业研报** **企业研报** 智能Agent系统。

02 要点

检索与采集

对接权威渠道，实时检索、采集主题相关的最新信息，作为生成研报的基础。

分析与生成

结合专业分析方法，对采集到的信息/数据进行梳理、归纳，生成研报的主体内容。

结构化与图表

对引用的数据进行结构转换，给段落生成必要的图表，提升研报的易读性。

03 探索

v0.1 Multi-Agent

定义管理智能体和成员智能体，通过智能体间的**多轮会话**与**Hand offs**完成任务。

01

问题:

- Token消耗巨大且不可控
- 生成研报质量不稳定
- 多轮会话经常出现死循环

02

v0.2 Rerank RAG

采集所有信息，分割→词嵌入→保存。再进行：检索→Rerank→生成，完成任务。

问题:

- 分割导致信息损失
- 检索信息缺失
- 信息库构建耗时较长
- 资源消耗大

v1.0 Data Menu RAG

搜索结果构建数据字典，词嵌入→保存。再通过：检索→详情采集→摘要→生成，完成任务。

03

优势:

- 无分割
- 模拟人类习惯
- 速度较快
- token消耗较小

03 解决方案

「 基于改进型RAG技术的金融研报生成智能体 」

01 背景

传统RAG主要运用场景

信息延迟: 离线

信息时效: 静态

信息规模: 大规模



金融研报场景

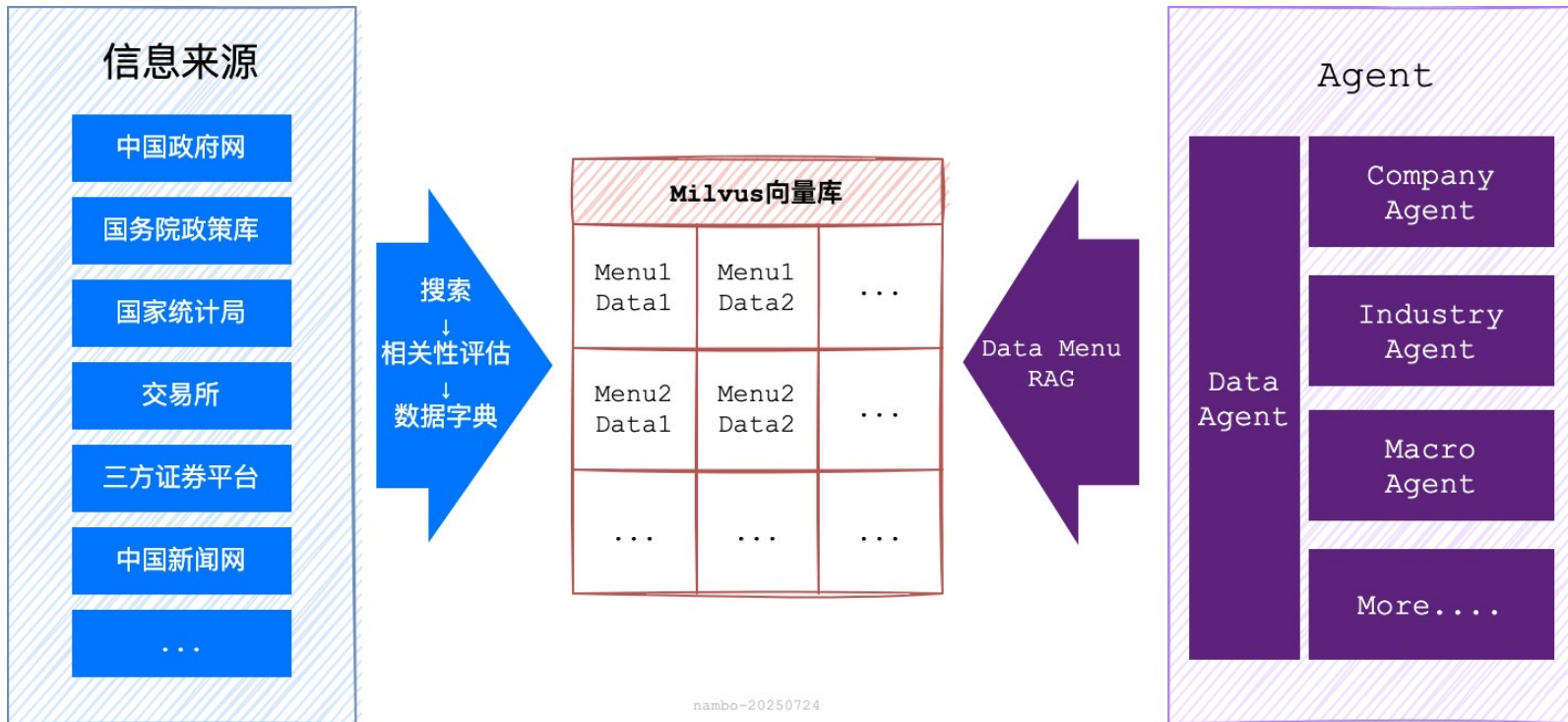
信息延迟: 在线

信息时效: 动态

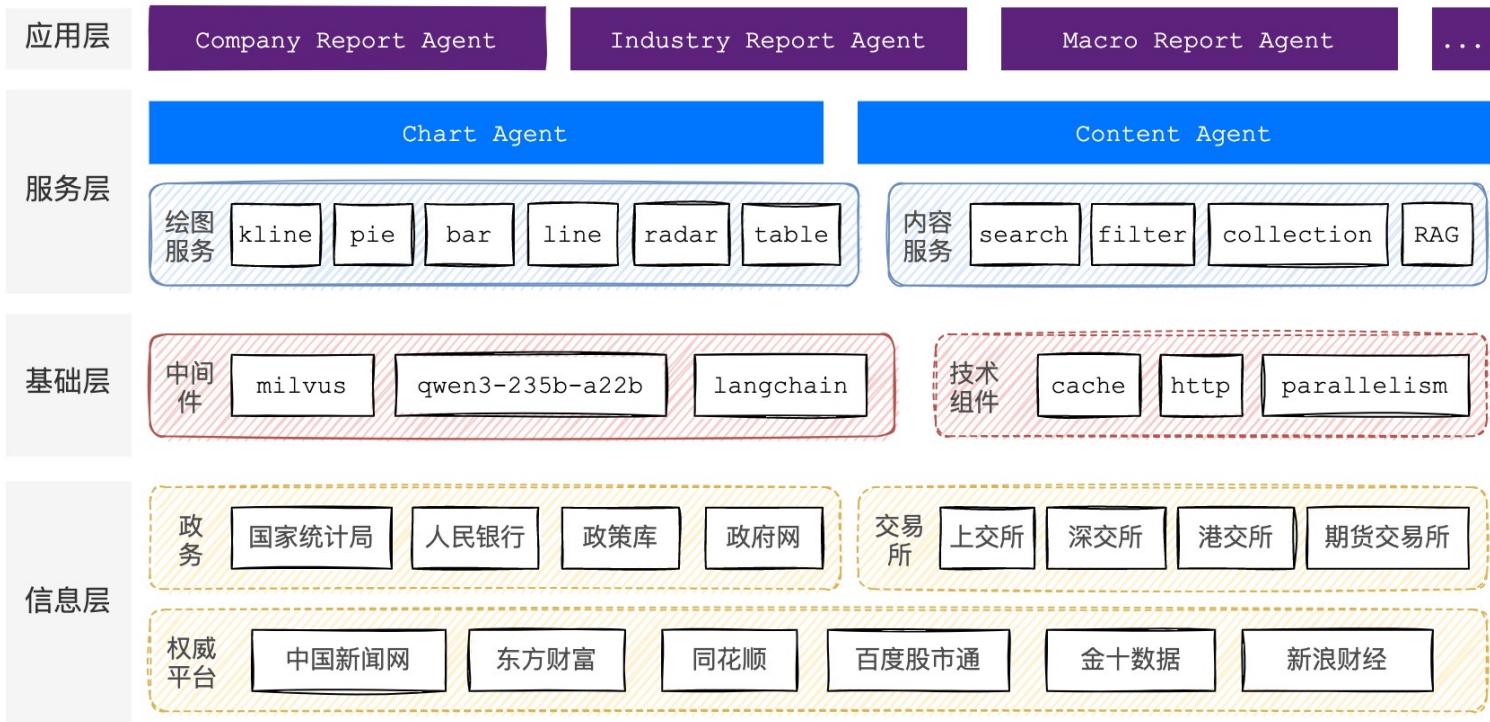
信息规模: 中等规模

根据生成公开信息深度研究报告场景的需求, 对传统RAG技术进行改进
实现了“**基于数据字典的RAG (Data Menu RAG)**”方案

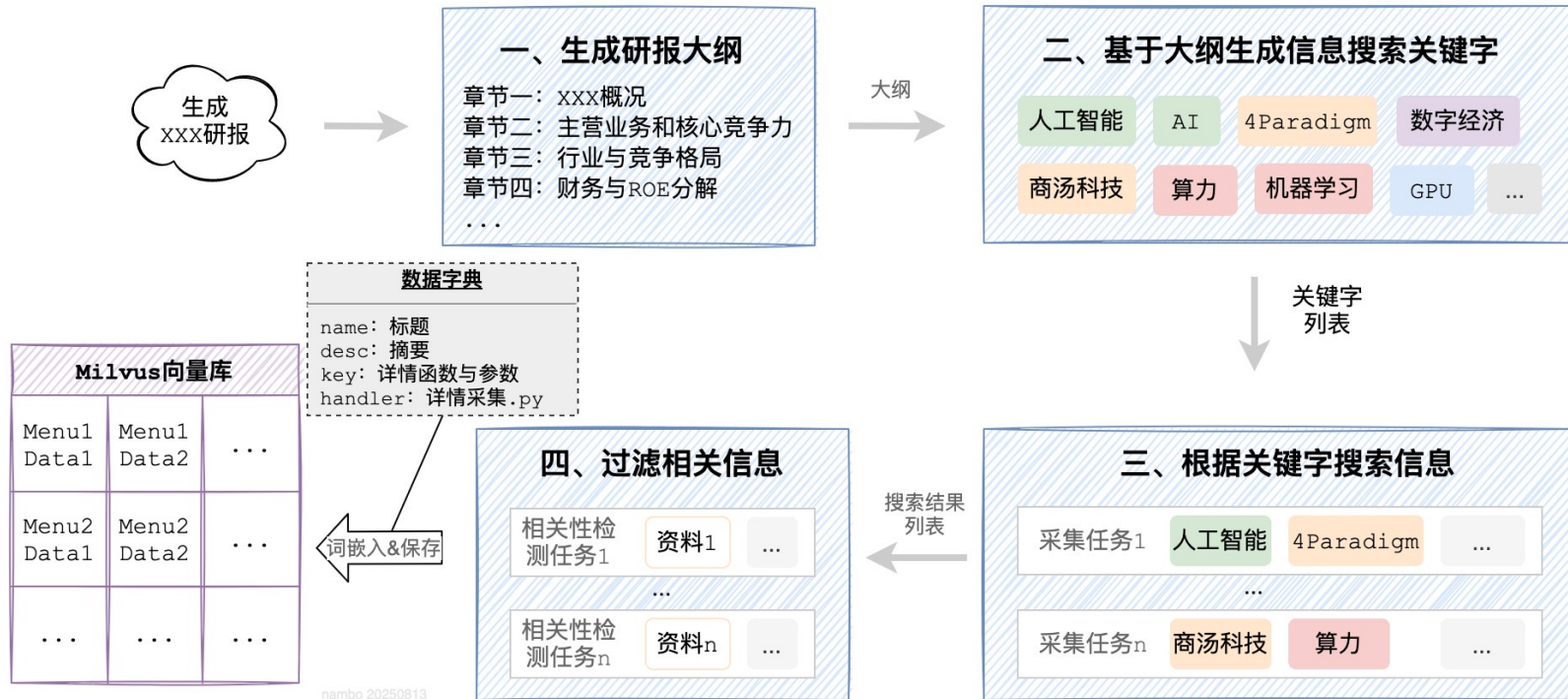
02 架构概览



03 应用架构



04 信息库构建过程



04 信息库构建过程

```
{  
  "name": "近13个月生产资料工业生产者出厂价格指数 (上月=100)",  
  "en_name": "A010808",  
  "url":  
    "https://data.stats.gov.cn/easyquery.htm?zb=A010808",  
  "desc": "来自中国国家统计局, 近13个月生产资料工业生产者出厂价格指数(上月=100), 所属分类: 工业生产者出厂价格分类指数, 包含字段: 时间、生产资料工业生产者出厂价格指数(上月=100)、采掘业生产资料工业生产者出厂价格指数(上月=100)、原料业生产资料工业生产者出厂价格指数(上月=100)、加工业生产资料工业生产者出厂价格指数(上月=100)",  
  "file_type": "csv",  
  "category": "工业生产者出厂价格分类指数",  
  "en_category": "",  
  "source": "中国国家统计局",  
  "date": "2025-07-01",  
  "key": "月度|A010808",  
  "handler": "data_gjtjj"  
}
```

数据的字典项示例

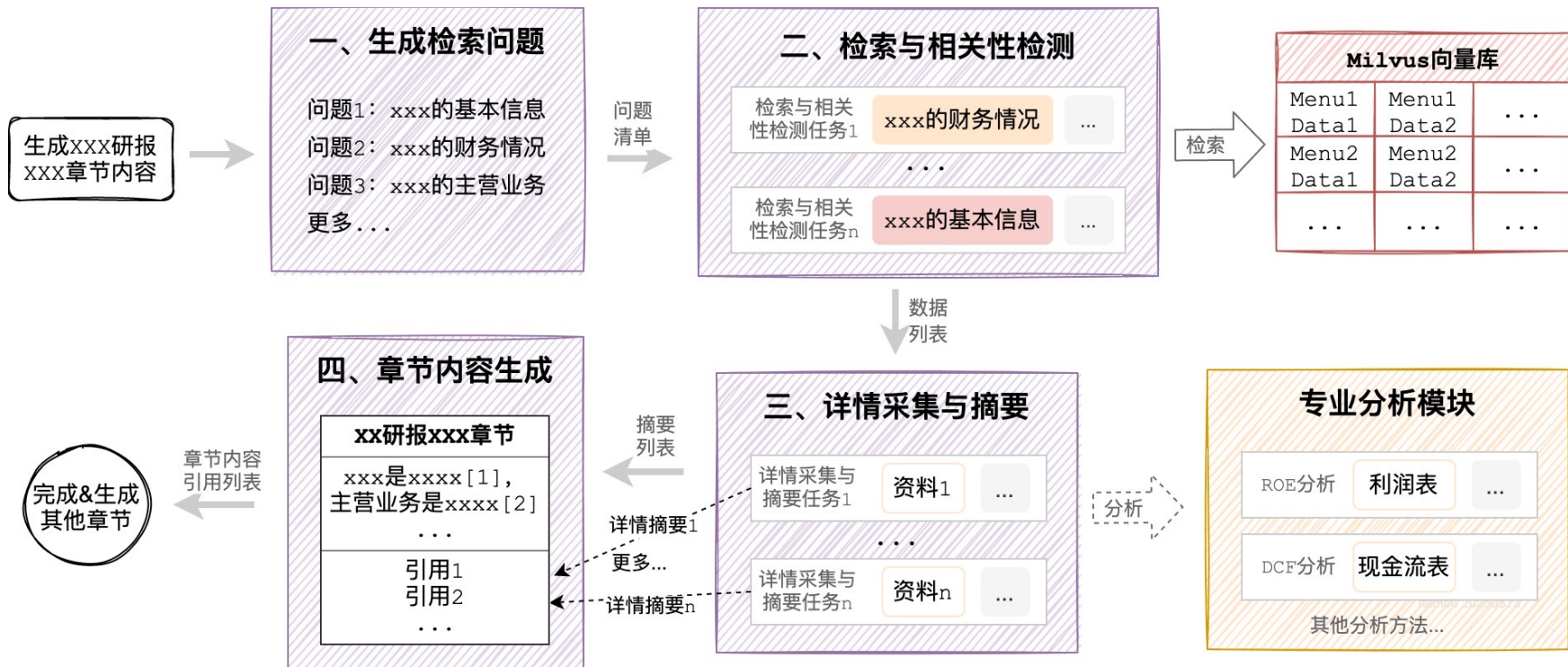
```
{  
  "name": "互联网数据耗尽? 世界人工智能大会业界热议“大模型养料”",  
  "en_name": "",  
  "url":  
    "http://www.chinanews.com.cn/cj/2025/07-28/10455102.shtml",  
  "desc": "上海商汤智能科技有限公司董事长兼CEO徐立指出, 互联网数据在最初的大模型发展中起到了非常关键的作用, 但如今正逐渐走向枯竭。他认为, AI想要获得进一步的突破性成长, 需要走出数字世界, 迈向物理世界, 强化大模型和具身智能的结合。",  
  "file_type": "htm",  
  "category": "临时:新闻-财经",  
  "en_category": "cj",  
  "source": "中国新闻网",  
  "date": "2025-07-28",  
  "key":  
    "http://www.chinanews.com.cn/cj/2025/07-28/10455102.shtml",  
  "handler": "news",  
  "relative": "是"  
}
```

文本的字典项示例

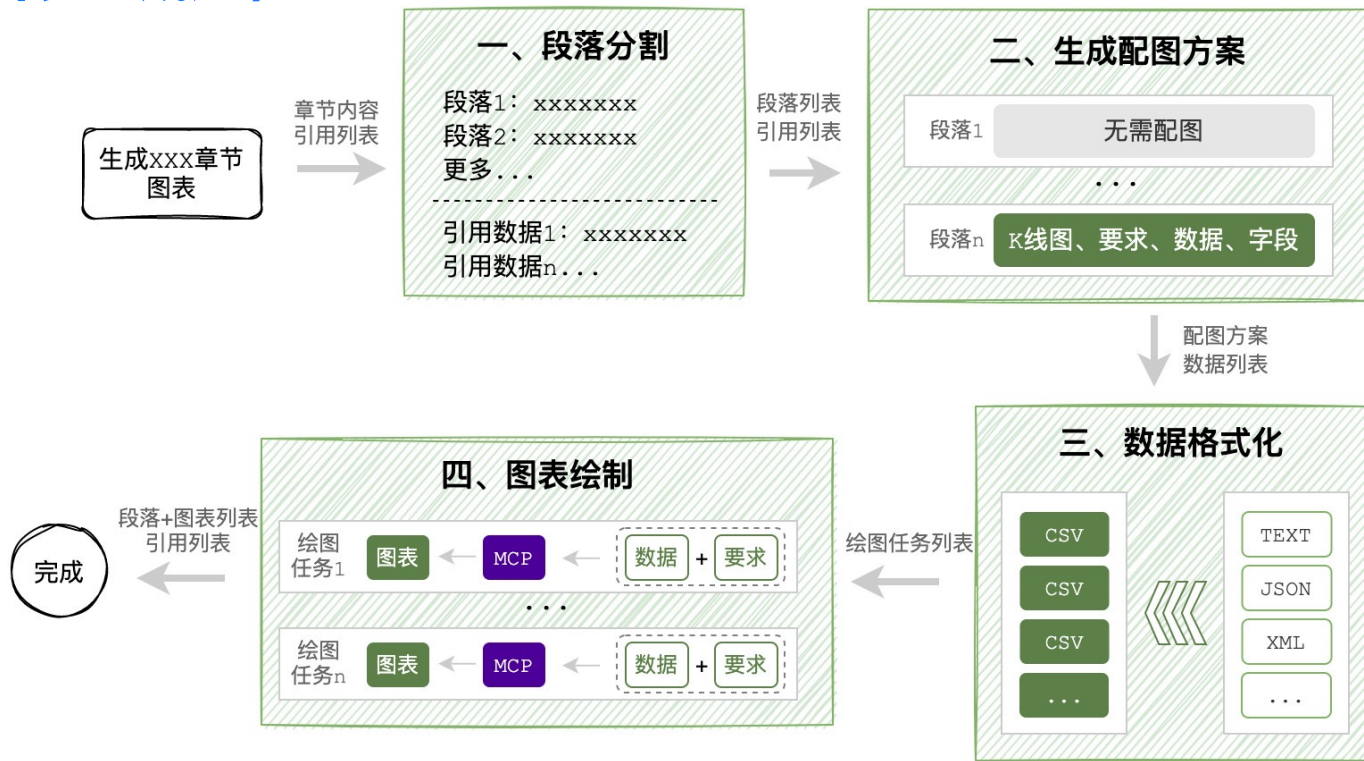
```
{  
  "name": "基于商汤科技财报进行的DCF+FCF估值分析",  
  "en_name": "DCF & FCF Valuation Report",  
  "url":  
    "https://www1.hkexnews.hk/listedco/listconews/sehk/2025/0424/2025042400918_c.pdf",  
  "desc": "上市公司商汤(股票代码00020.hk)基于现金流量表进行财务分析, 得到的现金流折现(DCF)与自由现金流(FCF)分析报告。包含历史现金流分析、未来现金流预测、折现率计算、企业估值及敏感性分析等内容。",  
  "file_type": "pdf",  
  "category": "临时:估值分析报告",  
  "en_category": "Valuation Analysis Reports",  
  "source": "香港证券交易所",  
  "date": "2025-07-26",  
  "key": "get_dcf_fcf_analysis|商汤科技|00020|hk",  
  "handler": "financial_analysis"  
}
```

自定义分析的字典项示例

05 内容生成过程




06 图表生成过程





04 方案成效

「 基于改进型RAG技术的金融研报生成智能体 」



01 赛题排名

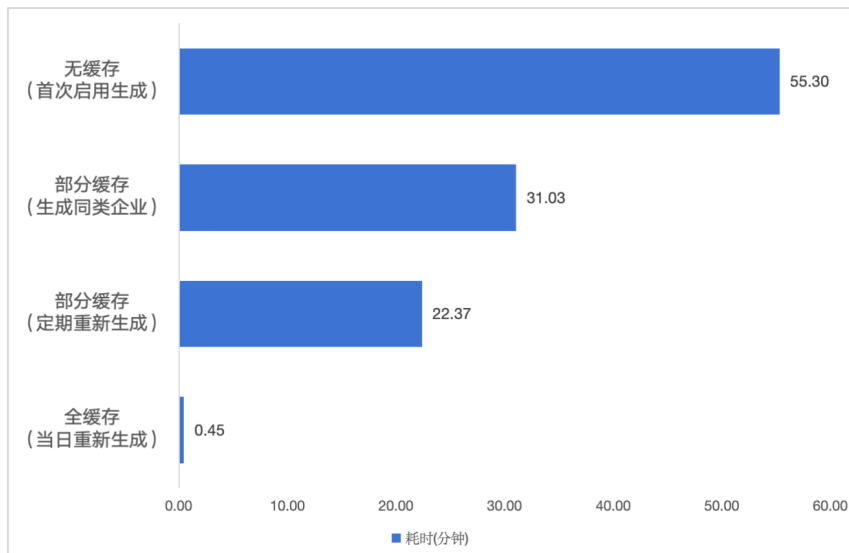
基于该方案，构建的最小可行版本智能体，在赛题中取得了**B榜综合排名第四**的成绩。

* 受限于时间和人力，未进行深入优化

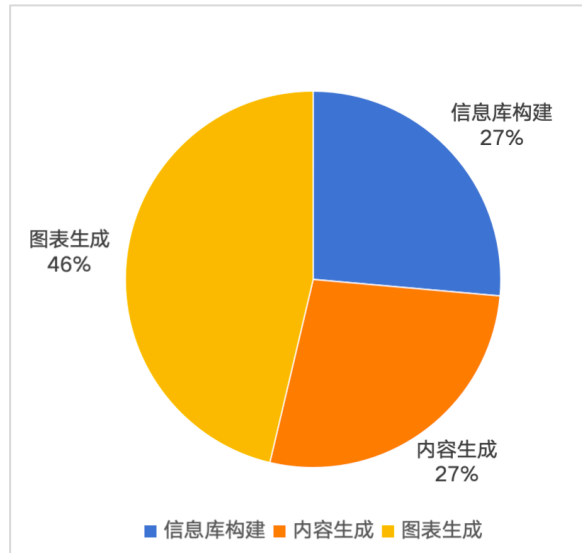
B榜	A榜	
排名	参与者	组织
1	RUC-NLPIR	中国人民大学
2	乘风破浪的弟弟	电子职业学院
3	清水河小分队	
4	队伍名字不能为空	CSRC BANK

02 执行速度

在命中缓存，并调高生成的并发度后，最快 **10分钟** 左右可生成一篇研报。



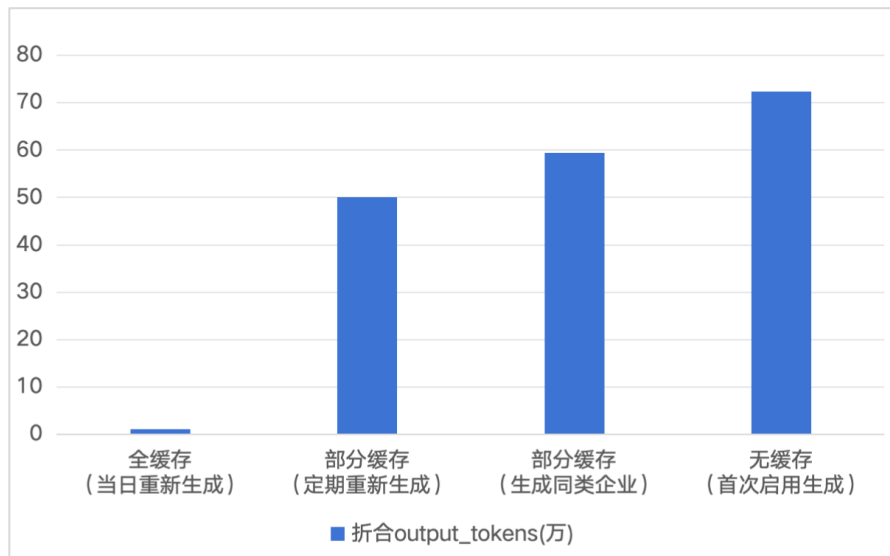
不同场景研报 (企业) 生成速度



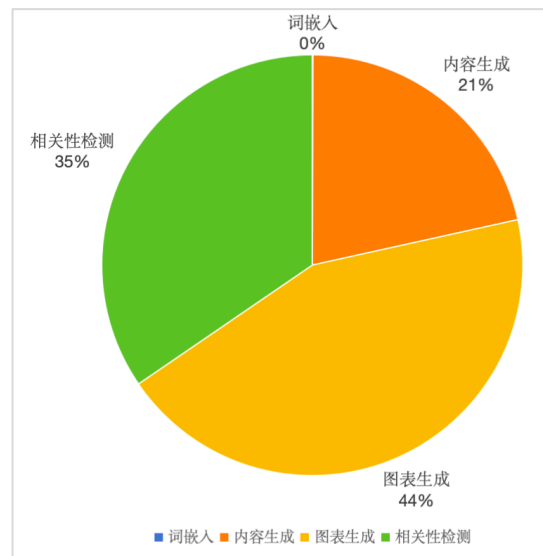
不同步骤耗时

03 资源消耗

在缓存命中率高的情况下，生成一篇研报，最少仅需消耗**30万token**左右。



不同场景token消耗对比



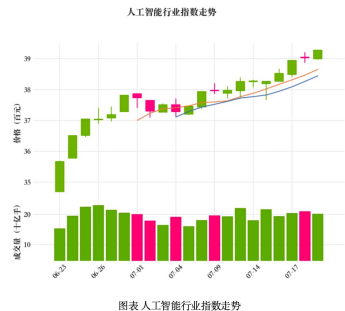
token消耗分布

04 生成效果

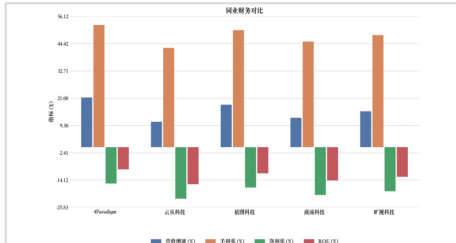
目录

一、公司概况	1
二、主营业务与核心竞争力分析	4
三、行业与竞争格局分析	7
四、财务报表分析	10
五、财务比率分析与 ROE 分解	13
六、股权结构与治理分析	16
七、估值模型与财务预测	19
八、投资建议与风险提示	22

三、行业与竞争分析



人工智能行业在 2025 年持续保持强劲增长态势，行业指数从 2025 年 6 月 23 日的 3494.21 点上涨至 7 月 21 日的 3927.35 点，累计涨幅达 12.4%，期间最高达 3927.35 点，成交量和成交额也呈现稳步上升趋势，7 月 21 日成交额达到 3192.49 亿元，显示出市场对人工智能行业的高度关注和资金活跃度^[1]，这一增长趋势表明人工智能产业正处于快速发展阶段，技术创新和应用场景的拓展推动了行业的持续扩张。



选取 5 家与 4Paradigm 业务相近的 AI 企业进行财务对比分析: 商汤科技、旷视科技、云从科技、依图科技和 4Paradigm。4Paradigm 以 21.3% 的营收增速领先同业，显示其在行业中的增长动能强劲，依图科技以 18.2% 增速次之，旷视科技 15.4%，商汤科技 12.6%，云从科技 10.9%。毛利率方面，4Paradigm 以 52.4% 居首，显示其产品结构和技术壁垒较强，依图科技 50.2%，旷视科技 48.1%，商汤科技 45.3%，云从科技 42.6%。盈利能力方面，所有公司均未实现盈利，净利率均为负值。4Paradigm 净利率-15.6%，优于同业平均水平，反映其成本控制 and 运营效率较好。ROE 方面，4Paradigm ROE 为-9.5%，优于同业平均水平，显示其资本使用效率相对较高^[2]。

参考文献

- [1] 中国新闻网. 推动人工智能产业迈向更高水平 [EB/OL]. (2025-07-23)[2025-07-23]. <http://www.chinanews.com.cn/gn/2025/05-05/10410876.shtml>
- [2] 中国新闻网. 报告: 2024 年中国 AI 产业规模突破 7000 亿元 [EB/OL]. (2025-07-23)[2025-07-23]. <http://www.chinanews.com.cn/cj/2025/04-26/10406330.shtml>
- [3] 中国新闻网. 发展不平衡不充分与因地制宜发展新质生产力 [EB/OL]. (2025-07-23)[2025-07-23]. <http://www.chinanews.com.cn/ll/2025/07-23/10452307.shtml>
- [4] 中国新闻网. 大力推动我国人工智能大模型发展 [EB/OL]. (2025-07-23)[2025-07-23]. <http://www.chinanews.com.cn/ll/2025/05-08/10412157.shtml>
- [5] 香港恒生指数有限公司. 恒生人工智能主题指数 [EB/OL]. (2025-07-23)[2025-07-23]. <https://www.hsi.com.hk/schi/indexes/all-indexes/hsait>
- [6] 中国人民政府网. 以“高效办成一件事”常态化推进为牵引，持续完善数字时代政府治理体系，提升治理水平 [EB/OL]. (2025-07-23)[2025-07-23]. https://www.gov.cn/jzhengce/2025/07/content_7031206.htm
- [7] 中国新闻网. 面对不确定性，小微企业为何强势回升? | 国是访问 [EB/OL]. (2025-07-23)[2025-07-23]. <http://www.chinanews.com.cn/gszc/2025/05-13/10414435.shtml>

研报章节

段落与复杂图表

专业 (财务) 分析

参考文献与引用标注

05 技术创新

基于数据字典的增强检索（Data Menu RAG）技术

01 基于数据字典的增强检索生成 (Data Menu RAG) 技术

创新的提出“基于数据字典的增强检索”技术，解决了传统RAG技术，在公开信息研报场景，存在的速度慢、资源消耗高、信息损失等问题，研报生成效率和质量大幅提升。

02 多级缓存与“断点重试”

免费数据源稳定性较差，在大模型会话、数据源获取等关键点设置了多层级的缓存。出现异常时，会自动清理异常缓存，回滚至最近成功点重新继续执行，确保100%成功生成研报，且单篇研报最低仅需消耗30万Token左右。

03 并行生成

在内容数据采集、图表绘制等耗时大的节点，实现了多任务并行生成。在调高并行度且命中大量缓存的情况下，最快10分钟左右可完成研报生成。

04 数据易接入

开发了一套数据接入规范，按照约定的数据字典结构，实现2个标准函数（出入参格式、函数名），即可快速接入新的数据源，用于研报生成。

05 场景可拓展

采用多层agent架构，服务层Agent已封装通用基础能力，通过应用层Agent简单的提示词调整，即可灵活、便捷的实现新场景拓展。

06 待改进点

「 基于改进型RAG技术的金融研报生成智能体 」

1

大模型

- 使用最新的模型，如：qwen3-235b-a22b-thinking-2507
- 开启大模型的深度思考/长思考

2

workflow

- 优化检测反馈机制（提高信息召回率、图文相关性、引用可靠性）
- 优化图表绘制（解决反复失败引起的耗时长、token消耗大等问题）

3

信息源

- 使用Tushare、恒生聚源等付费数据源，提升稳定性、全面性
- 接入更多行业（信通院、商联等）、证券（纽交所、伦敦等）数据

4

其他

- 完善信息隔离机制，支持同时生成多篇研报
- 提示词优化
- 提取Data Menu RAG相关方法，构建公共包
- 更多...



谢谢

队伍名字不能为空 · 储晓宇（常熟银行）

赛题四：智能体赋能的金融多模态报告自动化生成

