

2026 年 8 月 13 日



总量研究

“金融图书馆”一站式服务

——大语言模型结合 RAG 范式与 MCP 接口的运用实践

要点

核心观点：

本研究构建了一个面向金融领域的多源异构知识管理与智能分析系统。其核心贡献在于提出并验证了一套将非结构化金融文档转化为可检索、可交叉验证、可追溯的 LLM 决策支撑体系的工程范式。

系统的核心挑战是金融知识的异质性处理。书籍提供长期的理论框架但缺乏时效性，广义研究报告提供最新观点但置信度参差不齐，数据商提供精确的量化数据但缺乏语义解释。这三种知识形态在格式、时效、确定性三个维度上存在本质差异。本系统的关键设计在于：不是简单地对文档做 RAG 检索然后丢给 LLM，而是为每类信息建立了不同的处理流水线 and 置信度标注机制。

研报分析的五阶段流水线是该范式的代表性实现。其方法论价值在于它展示了一套可泛化的 LLM 质量保障机制：观点提取→独立检索→证据匹配→事实核查→非对称置信度评估。其中核查环节将 LLM 的语义理解与确定性数值比对相结合，为非结构化分析提供了客观锚点。

可追溯性是本系统区别于通用 LLM 的根本特征。从金融问答到资产配置展望，系统的每一次输出都携带完整的来源映射。LLM 生成的每项资产评分都可以回溯到具体的数据点、研报条目或内部知识库片段。

风险提示：

本文仅为 AI 技术层面的试验观察、不构成任何投资建议。

作者

分析师：王开

执业证书编号：S0930526050001
021-52523416
wangkai9@ebsecn.com

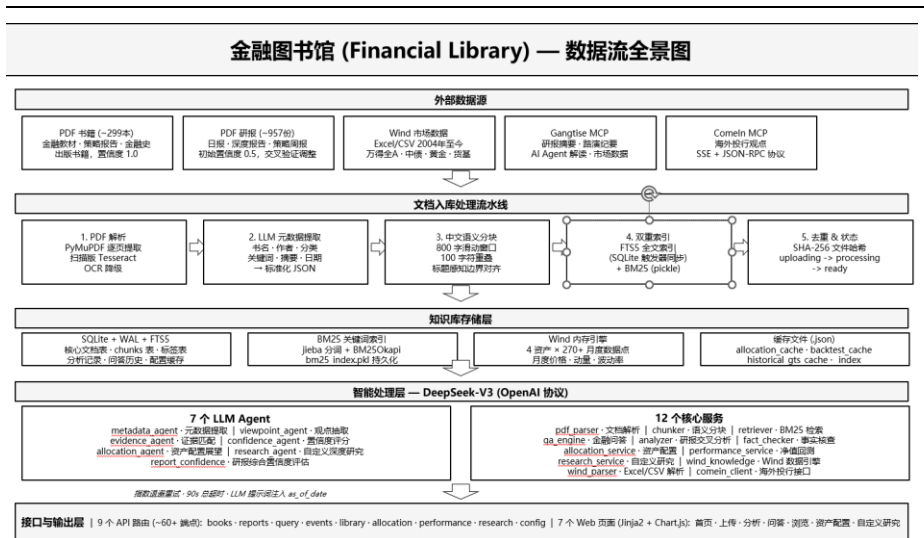
目录

一、实验框架：LLM 资产配置研究体系.....	3
二、系统架构与基础设施.....	4
2.1 系统架构介绍	4
2.2 启动流程	4
2.3 数据层管理	4
2.4 服务层介绍	5
三、Agent 与核心 workflow	6
3.1 Agent 体系设计.....	6
3.2 workflow 详解	7
3.3 Allocation Agent.....	7
四、接口与呈现层	8
4.1 API 总览与关键 API 端点	8
4.2 前端界面.....	8
五、外部集成与部署	9
5.1 MCP 调用	9
5.2 部署与运维	9
六、资产配置与回测系统	10
6.1 多元数据处理	10
6.2 历史回测引擎	10
6.3 时点回测	11
七、风险提示	11

一、实验框架：LLM 资产配置研究体系

我们所构建的“金融图书馆”是一个本地运行的 AI 驱动金融知识工作站。系统将 PDF 书籍、研究报告、市场数据、外部研报平台和海外智库观点整合为统一知识库，由 DeepSeek-V3 大模型驱动，提供文档入库与检索、研报分析、金融问答、大类资产配置、历史回测和时点回测六大核心能力。其根本设计目标不是替代量化模型或人类分析师，而是为金融从业者构建一个可检索、可交叉验证、可追溯决策链条的 AI 辅助分析环境。

图 1：数据流全景图



资料来源：光大证券研究所。

图 2：功能矩阵

模块	核心能力	输入	输出
文档入库	PDF 解析 + 语义分块 + BM25/FTS5 双索引	PDF 文件	结构化文档库
智能检索	jieba + BM25 + 时效性衰减 + 置信度加权	关键词/问题	Top-K 文档块 (含来源标注)
金融问答	检索增强 LLM 生成 (书籍+研报溯源)	自然语言问题	带引用来源的专业回答
研报分析	5 阶段流水线 + Wind 事实核查	一份研报	观点级置信度 + 矛盾分析 + 综合评估
资产配置	多源数据并行采集 → LLM 多维度评分合成	数据源开关 + 可选时点	周/月/季三维展望 + 风格二层拆分
回测体系	LLM 引导历史评分 + 时点回测 (无前视偏差)	回看年数 / 历史日期	净值曲线 + 绩效指标 + 排名对比

资料来源：光大证券研究所。

图 3：技术栈

层级	技术选型
Web 框架	FastAPI + Uvicorn (async)
数据库	SQLite + SQLAlchemy 2.0 (async) + WAL 模式 + FTS5
LLM	DeepSeek-V3 (OpenAI 兼容协议, openai SDK 调用)
中文处理	jieba 分词 (搜索引擎模式) + rank-bm25 (BM25Okapi)
数据处理	Pandas + NumPy + openpyxl
PDF 解析	PyMuPDF (fitz) + 可选 OCR 降级 (pytesseract)
前端	Jinja2 服务端渲染 + Chart.js 4.4 可视化
数据验证	Pydantic v2 + pydantic-settings
外部集成	MCP (subprocess CLI) + MCP (SSE/JSON-RPC)

资料来源：光大证券研究所。

二、系统架构与基础设施

2.1 系统架构介绍

系统采用五层架构：前端层 (Jinja2 + Chart.js) → API 层 (9 个 FastAPI 路由模块) → 服务层 (文档处理 / 资产配置两大集群) → Agent 层 (7 个 LLM Agent) → 数据层 (SQLite + BM25 + Wind + 外部数据源)。每层职责独立，通过 HTTP 和内部函数调用通信。设计原则是让 LLM 只出现在 Agent 层，其余层次保持确定性行为，以此控制系统的可调试性和可审计性。

2.2 启动流程

FastAPI lifespan 管理启动和关闭两个阶段。启动阶段执行五项初始化：(1) 强制设置标准输出为 UTF-8 (解决 Windows 环境下 GBK 编码导致的日志乱码问题)；(2) 创建 metadata/extracted/chunks/books/reports/ 等必要目录；(3) 初始化 SQLite 数据库，开启 WAL 模式和 FTS5 全文索引；(4) 后台异步构建 BM25 索引，不阻塞 HTTP 服务启动；(5) 将 Wind 市场 CSV 数据加载到内存对象 (4 资产 × 270+ 月度数据点)。关闭阶段取消后台任务并关闭数据库连接池。

2.3 数据层管理

系统使用 SQLite 作为唯一数据库，核心决策是归一化设计：所有文档类型 (书籍、研究报告、外部来源) 共享同一张 documents 表，通过 doc_type 和 source 字段区分来源，以 source_institution 字段标注归属机构。归一化设计的代价是部分字段 (如书籍的 ISBN、研报的分析师署名) 对特定类型留空，但收益是统一的查询和检索接口，以及新来源类型零成本扩展。

图 4：数据库表设计

表名	用途	关键设计
Documents	核心文档表，归一化存储所有来源	doc_type(book/report/external), file_hash(SHA256 去重), confidence(书籍 1.0/研报 0.5-0.85), status 状态机
chunks	文本块表，语义分块结果	document_id(FK CASCADE), chunk_index, text, token_count, section_title
tags	标签关联表	document_id(FK), tag_name, tag_category
analysis_records	研报分析结果 (JSON)	report_id(FK), analysis_result, supporting/contradicting_sources, document_confidence
qa_history	问答历史	question, answer, source_chunks(JSON), answer_confidence
chunks_fts	FTS5 虚拟全文索引	触发器自动同步 INSERT/DELETE/UPDATE, 支持中文模糊搜索

资料来源：光大证券研究所。

关键工程决策：(1) WAL 模式使读操作永不阻塞写操作，适合高频查询 + 低频写入的使用模式；(2) 外键级联删除 (ON DELETE CASCADE) 确保删除文档时自动清理关联的 chunks、tags 和 analysis_records；(3) 文件哈希去重防止重复导入同一 PDF；(4) 灵活字段 (authors、keywords、analysis_result) 以 JSON 文本存储，避免频繁 ALTER TABLE。

BM25 关键词检索链路：SQLite chunks 表 → jieba 分词 (搜索引擎模式) → BM25Okapi 向量化 → pickle 序列化至 data/bm25_index.pkl。最终评分公式为 $\text{FinalScore} = 0.5 \times \text{BM25_norm} + 0.5 \times \text{TimeDecay} \times \text{Confidence}$ ，其中 $\text{TimeDecay} = \exp(-0.001 \times \text{days_elapsed})$ 使新文档获得更高权重，Confidence 区分书籍 (权威性 1.0) 和研报 (时效性强但置信度 0.5-0.85)。持久化策略：启动时比对 chunk 总数，匹配则加载 pickle (秒级恢复)，不匹配则全量重建 (分钟级)。

选择 BM25 而非向量嵌入的原因是务实的：中文金融术语 (如信用利差久期管理) 对语义向量的敏感性较低，而关键词匹配的精确性和可解释性更适合本系统的可追溯设计原则。混合检索 (BM25 关键词 + FTS5 全文) 足以覆盖当前使用场景，且无需 GPU 或外部向量数据库。

WindKnowledgeBase 以内存对象形式维护四类资产的完整月度价格时间序列 (万得全 A / 中债综合指数 / 黄金 Au9999 / 货币基金指数，覆盖 2004-01 至 2026-07)。核心 API 包括：get_price() 取最近交易日价格，compute_return() 计算指定期间收益率 (自动对齐月末日期)，get_stats() 输出均值 / 标准差 / 最大回撤 / 多周期动量，以及 get_wind_snapshot(as_of_date) 获取截至任意历史日期的市场快照，后者是时点回测的关键依赖。所有接口均支持日期截断，确保回测中不会泄漏未来信息。

2.4 服务层介绍

服务层包含 12 个模块，按功能分为两大集群：文档处理集群 (pdf_parser → chunker → retriever → qa_engine → analyzer → fact_checker → wind_knowledge → wind_parser) 和资产配置集群 (allocation_service → performance_service → summary_client → research_service)。

图 5：服务层文档处理集群介绍

服务文件	核心类/函数	职责	关键参数/设计
pdf_parser.py	parse_pdf(), extract_text()	PyMuPDF 提取 PDF 文本，计算文件哈希	支持扫描件 OCR 降级 (pytesseract)
chunker.py	SemanticChunker.chunk()	语义分块，800 字符窗口 + 100 重叠	中文标点边界检测，最小块 100 字符
retriever.py	HybridRetriever	jieba + BM25 混合检索	beta=0.5(BM25) + gamma=0.5(时效性×置信度)
qa_engine.py	FinancialQAEngine.answer()	检索增强 LLM 问答	最多 6000 字符上下文，每文档最多 2 chunks, 总超时 90s
analyzer.py	ReportAnalyzer.analyze()	5 阶段研报分析流水线编排	每观点独立检索，非对称惩罚，Wind 核查
fact_checker.py	FactChecker.check()	LLM 提取量化论断 → Wind 确定性比对	精确比对±3%，近似比对±15%，阈值比对
wind_knowledge.py	WindKnowledgeBase	Wind 内存查询引擎	价格/收益/统计/快照，支持日期截断
wind_parser.py	parse_wind_csv()	Wind Excel/CSV 解析 → 标准化格式	CSV 路径 → DataFrame

资料来源：光大证券研究所。

三、Agent 与核心 workflow

3.1 Agent 体系设计

系统共定义 7 个 LLM Agent, 每个 Agent 封装了特定的 System Prompt、JSON Schema 输出约束和响应解析逻辑。Agent 层的设计原则是：(1) 每个 Agent 只承担单一明确的任务，通过组合而非修改来扩展能力；(2) Prompt 中显式标注角色设定和输出格式，降低 LLM 生成偏差；(3) 所有 Agent 输出为结构化 JSON，上层服务负责校验和编排。

图 6：Agent 清单

Agent 文件	核心类/函数	LLM 角色	输入	输出	所在 workflow
viewpoint_agent.py	ViewpointExtractor	金融研报分析师	PDF 文本	结构化观点列表 (JSON)	研报分析 Phase 1
evidence_agent.py	EvidenceMatcher	证据审查员	观点 + BM25 上下文	supporting/contradicting/neutral	研报分析 Phase 3
confidence_agent.py	ConfidenceAssessor	置信度评估师	观点 + 证据 + Wind 核查	0-1 置信度分数	研报分析 Phase 4
metadata_agent.py	MetadataExtractor	元数据提取器	PDF 文本首部	title/authors/category/keywords/abstract	文档入库
report_confidence.py	ReportConfidence	文档级置信度评分	全部观点置信度 + 矛盾	文档级 0-1 评分	研报分析 Phase 5
allocation_agent.py	AllocationOutlook	首席大类资产配置策略师	Wind 快照 + 摘要 + 内部库	资产排序+风格+行业+主题	资产配置展望
research_agent.py	ResearchReport	深度研究员	自定义主题 + 多源检索	结构化研究报告	自定义研究

资料来源：光大证券研究所。

3.2 workflows 详解

研报分析是本系统最具方法论特色的 workflow。一份已入库的研报经过五个阶段逐层处理：

Phase 1 — 观点提取：ViewpointExtractor 逐段提取投资观点，输出标准化 JSON，包含 claim（核心论断）、asset_category（资产类别）、direction（看多/看空/中性）、timeframe（时间维度）和 rationale（推理依据）。

Phase 2 — 上下文检索：每个观点独立进行 jieba 分词 + BM25 检索 Top-15 相关块，同时进行矛盾定向检索，对关键词做反转（如看多→看空），搜索与当前观点相反的历史论述。

Phase 3 — 证据匹配：EvidenceMatcher 对每个观点及其上下文块进行证据归类，输出 supporting（支持该观点的原文片段）、contradicting（矛盾的原文片段）和 neutral（相关但不表态）三类标注。

Phase 3.5 — Wind 事实核查：FactChecker 先从观点中提取可量化论断（如过去 3 个月上涨超 15%），再用 Wind 数据进行确定性数学比对。精确比对允许 $\pm 3\%$ 偏差，近似比对允许 $\pm 15\%$ ，超出阈值则为无法验证。输出三种判定：confirmed / contradicted / unverifiable。

Phase 4 — 置信度评估：ConfidenceAssessor 综合观点本身的论证质量、证据匹配结果和 Wind 核查结论，输出 0-1 分数。关键设计是非对称惩罚：矛盾证据最高扣 0.35 分，而支持证据最高仅加 0.15 分。这里的逻辑是：一条矛盾证据的破坏力大于一条支持证据的建设力，因为金融分析中的看似合理远比明显错误更危险。

Phase 5 — 矛盾综合：按文档分组聚合所有观点的置信度，区分同期分歧（同一报告内部自相矛盾）和事后证伪（观点被其他来源或后来数据否定），最终输出文档级置信度评分和综合评估报告。

这套流水线的方法论价值不限于分析研报这一具体任务。观点提取→独立检索→证据匹配→事实核查→非对称置信度的五阶段框架可泛化至任何需要 LLM 做财经文本分析且要求结果可验证的场景。

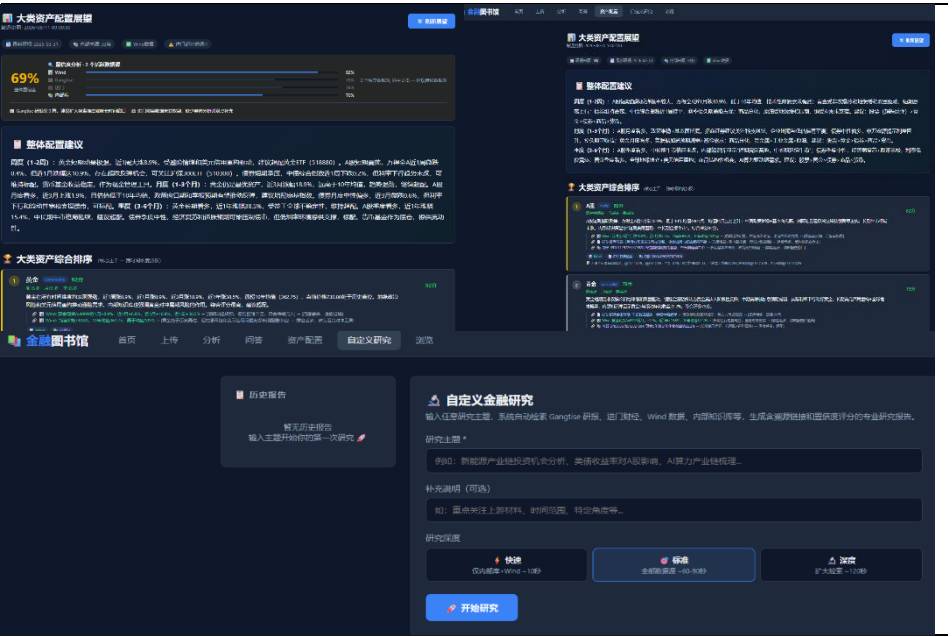
3.3 Allocation Agent

AllocationOutlook 是系统中 Prompt 设计最复杂的 Agent。报告采用主干→支线逻辑架构：主干是基于全部数据源的大类资产排序，支线依次为风格偏好、行业配置和主题热点。

三个时间维度的 bias 设计使得同一资产在不同持有周期下可以有不同的方向判断，例如黄金可能在周度看多（动量驱动）但季度看空（估值均值回归驱动）。这种多维度的输出格式比单一评分更能反映真实的投资决策复杂性。

评分公式
$$\text{composite_score} = \text{evidence_strength} \times 0.5 + \text{cross_validation} \times 0.4 + \log_2(\text{source_count} + 1) \times 0.1$$
 中，evidence_strength 侧重 Wind 可量化数据的支撑程度，cross_validation 衡量和内部库的来源共识度，source_count 对数衰减版本确保边际信息来源的贡献递减。权重分配体现了一个判断：在资产配置中，硬数据的支撑（50%）和多元来源的交叉验证（40%）远比信息来源的数量（10%）重要。

图 7：Allocation 界面展示



资料来源：光大证券研究所。

四、接口与呈现层

4.1 API 总览与关键 API 端点

9 个 FastAPI 路由模块，47 个 RESTful 端点，按功能域组织：

图 8：API 路由全景

路由模块	前缀	端点数	核心职责
books	/api/books	7	书籍 CRUD + 上传 + 处理 + 批量扫描
reports	/api/reports	12	研报 CRUD + 上传 + 批量扫描 + 分析
query	/api/query	4	金融问答 + 关键词搜索 + 状态 + 历史
library	/api/library	7	统一搜索 + 标签统计 + 分块/全文/PDF 预览
events	/api/events	2	宏观事件 CRUD
allocation	/api/allocation	4	生成展望 + 读取缓存 + 状态 + QA
performance	/api/performance	4	回测 + 时点回测 + 读取缓存 + 清除缓存
research	/api/research	4	自定义研究生成 + 列表 + 获取 + 删除
config_api	/api	3	Wind 资产列表 + 配置查看 + 配置更新

资料来源：光大证券研究所。

4.2 前端界面

前端采用 Jinja2 服务端渲染，base.html 提供全局布局（CSS 自定义属性

实现暗色主题）、Chart.js 4.4 CDN 引入和导航栏。七个页面对应七个功能入口。选择 SSR 而非 SPA 的原因是务实的：项目面向单用户本地使用，无需前端构建工具链，Jinja2 模板直接渲染后端数据，开发路径最短。

图 9：前端界面路由

路由	模板	功能
/	index.html	首页仪表盘：统计卡片 + 分类分布 + 最近文档
/upload	upload.html	文档上传：单文件拖拽 + 批量目录扫描
/analyze	analyze.html	研报分析：选择 → 触发 5 阶段流水线 → 逐观点展示
/qa	qa.html	金融问答：聊天式交互 + 引用溯源 + 历史记录
/library	library.html	图书馆浏览：搜索 + 标签筛选 + 分块 + PDF 预览
/allocation	allocation.html	资产配置主页：展望 + 回测 + 时点回测（最复杂页面）
/research	research.html	深度研究：自定义主题 → 结构化报告

资料来源：光大证券研究所

allocation.html 是前端最复杂的页面，包含三个功能面板：配置展望（资产排名卡片 + LLM 摘要 + 风格拆分柱状图 + 行业建议）、历史回测（Chart.js 累计净值双线图 + 月度权重 Stacked Area 图 + 四指标卡片 + 当期风格权重柱状图）、时点回测（日期选择器 + LLM 排名 vs 实际收益对比表 + Top1 命中标识）。所有面板通过 Fetch API 异步加载，含进度指示器。

五、外部集成与部署

5.1 MCP 调用

Gangtise（勾股）是研报数据平台，系统通过 subprocess CLI 方式调用其 Python 脚本。4 个技能包覆盖：(1) gangtise-file (14 scripts)，按关键词/日期/机构搜索并下载研报全文；(2) gangtise-agent (5 scripts)，AI 解读热门主题和股票线索；(3) gangtise-data (12 scripts)，行情/财务/估值/资金流/股东数据；(4) gangtise-kb (3 scripts)，知识库级别检索。调用方式为 subprocess.run()，stdout JSON 经正则解析提取。资产配置模块按 5 个维度并行搜索，单维度 90s 超时，最终按 file_id 去重合并。选择 subprocess CLI 而非 MCP SDK 的原因是团队提供的即为 Python 命令行脚本，直接调用省去了适配层，且稳定性可控。

知识库平台提供海外智库观点数据。集成协议为 SSE + JSON-RPC：GET SSE 端点获取 sessionId 和 messageUrl → POST JSON-RPC initialize → tools/list → tools/call。支持查询券商分析师观点、投研数据、指数行情、宏观指标、热门路演和 ETF 资金流。有两点限制：(1) 仅支持当期查询，历史时点回测时跳过此数据源；(2) SSE 连接需维持长连接，网络中断后需重建会话。

5.2 部署与运维

系统通过 .env 文件管理全部配置，必填项仅三项：DEEPSEEK_API_KEY（在 platform.deepseek.com 注册获取）、DEEPSEEK_BASE_URL（默认 https://api.deepseek.com/v1）和 LLM_MODEL（推荐 deepseek-chat）。平台的 GTS_ACCESS_KEY 和 GTS_SECRET_KEY 为选填项，缺省时资产配置模块自行降级，仅依赖 Wind 数据和内部知识库生成展望。启动命令为 python -m uvicorn src.main:app --reload --host 127.0.0.1 --port 8000，须在项目根目录执行而非 src/ 子目录，否则会因

Python 模块搜索路径问题报 `ModuleNotFoundError: No module named 'src'`。生产环境建议去掉 `--reload` 并增加 `--workers 4` 启用多进程。健康检查端点 `GET /health` 返回三项关键状态，BM25 索引是否就绪、已索引 chunk 数、Wind 数据加载状态和日期范围，无需额外配置即可用于监控探测。系统所有持久化数据集中在 `data/` 目录 (SQLite 主库、BM25 索引、各类缓存 JSON)，与 `books/` 和 `reports/` 中的原始 PDF 以及 `.env` 配置文件共同构成完整备份单元，四者齐备即可在新环境中恢复全部系统状态。从解压 ZIP 到浏览器打开首页，整个部署流程不超过五分钟，所有 Python 依赖声明在 `requirements.txt` 中，一行 `pip install -r requirements.txt` 完成环境准备。

六、资产配置与回测系统

6.1 多元数据处理

`generate_outlook()` 的并行采集策略是系统的性能瓶颈和设计重点。四个数据源通过 `asyncio.gather` 并发获取：(1) `fetch_gangtise_reports()` 按 5 个维度各启动一个 subprocess，调用 CLI 搜索研报，每维度 90s 超时，合并去重；(2) `fetch_summary_data()` 通过 SSE 连接 + JSON-RPC 获取海外分析师观点（仅当期）；(3) `search_internal_library()` 按 6 组 BM25 查询关键词并行检索内部知识库；(4) `get_wind_snapshot()` 获取四资产当前价格、1 周/1 月/3 月/1 年动量统计和 10 年期历史统计。四个数据源中任一失败不影响其余，系统以降级模式继续运行。

AllocationOutlook Agent 接收聚合后的结构化 Prompt 上下文，输出包含 4 项资产的排名 (`composite_score` 0-100)、三个时间维度的方向性 bias (周/月/季)、风格偏好 (成长/价值/大盘/小盘/高股息各自超配/标配/低配)、行业配置建议和主题分析。`_source_map` 字段记录了每项结论的数据来源，使资产评分可以追溯到具体的 Wind 数据点、报告或内部文档。

6.2 历史回测引擎

由于 2004 年的数据无 12 个月之前的参考，回测引擎在 2005-01 至 2026-07 的约 250 个逐月末日期上运行。每期执行：基于截至当月的数据计算评分 → 归一化为四资产权重 → 持有一个月 → 记录策略净值。基准为等权组合，即每月四资产各占 25%。

历史每期的评分采用五因子加性模型： $\text{score} = \text{base_bias} + \text{sentiment_adj} + \text{momentum_adj} + \text{value_adj} + \text{vol_penalty}$ 。

`base_bias`：将配置展望阶段 LLM 产出的月度 bias 文本映射为数值，看多/超配→80，偏多→68，中性→50，偏空→32，看空→20。这是评分的锚，后续四项在此基础上做 ± 调整。

`sentiment_adj` (±10)：读取该时期研报中的情感关键词（看多/上涨/增配/利好为正，看空/下跌/减配/利空为负），综合多篇取均值。

`momentum_adj` (±10)：近 3 个月资产收益率与该资产历史中位数的偏离，高于中位加分（动量延续假设），低于中位减分。

`value_adj` (±8)：当前价格与 12 个月移动均线偏离，价格低于均线加分（均值回归假设），高于均线减分。股票资产使用 $(MA-P)/MA \times 20$ 公式并限幅到 $[-8, +8]$ 。

vol_penalty (-6~0): 近 3 个月波动率与历史中位数的比对，波动率位于高分位扣 6 分，中分位扣 3 分，低分位不扣分。这是纯惩罚项，没有加分可能。

scores_to_weights() 将四项资产评分映射为组合权重，经过四步：(1) 将评分裁剪到 [0, 1] 区间；(2) 归一化 $w_i = s_i / \sum s_i$ ；(3) 下限保护 $w_i = \max(0.05, w_i)$ ，确保任何资产不低于 5%；(4) 重新归一化使权重之和为 1。下限保护的引入防止单一资产极度偏离时组合退化为单一资产策略。

6.3 时点回测

时点回测是本系统在方法论上最具创新性的组件。与传统回测的不同在于：它不是用历史数据跑一个策略看看收益，而是回到过去某个时间点，用当时可用的全部信息让 AI 独立做一次配置判断，然后看这个判断后来被市场证实还是证伪。

实现机制分为五阶段：(1) 数据截断，Wind 价格止于 as_of_date，研报以 as_of_date 为 end_date 上限，语料库 跳过（仅当期可用），LLM 提示词注入日期锚定 + 禁止引用未来数据；(2) 完整配置流程，调用 generate_outlook(as_of_date, force_refresh=True)，与真实使用时完全相同的流水线；(3) 评分 → 权重，通过 LLM_NAME_TO_CODE + CATEGORY_TO_CODE 双重降级桥接将 LLM 使用的资产名称映射到 Wind 代码，再计算权重；(4) 后续收益，使用月末对齐日期计算 1/3/6 个月的前向实际收益；(5) 排名对比，LLM 评分排名与后续实际收益排名制成对比表，标注 Top1 命中和偏差方向。

实测验证中，as_of_date=2024-12-31 时 LLM 判断黄金为 Top1 (92 分，看多)，3 个月后黄金实际上涨 +18.87% 排第一；as_of_date=2025-06-30 时 LLM 判断 A 股为 Top1，3 个月后 A 股上涨 +19.46% 排第一。两次验证均命中 Top1，初步证实了方法的可行性。

然而，必须指出时点回测的根本性限制：2005-2018 年期间由于 MCP 缺乏稳定的历史覆盖，评分引擎中 sentiment_adj 项的输入退化为空，使得这一时期的评分实质上趋近于纯量化规则驱动。这一信息基础的不对称性意味着全周期回测绩效不能被解释为 AI 配置能力的均质表现，不同时期 LLM Agent 可用的信息集存在结构性差异，在评估中必须予以区分。这一约束不仅限于本研究，而是所有依赖外部数据源的 AI 金融系统的共同挑战。

七、风险提示

本文仅为 AI 技术层面的试验观察、不构成任何投资建议。

行业及公司评级体系

评级	说明
买入	未来 6-12 个月的投资收益率领先市场基准指数 15%以上
增持	未来 6-12 个月的投资收益率领先市场基准指数 5%至 15%；
中性	未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差-5%至 5%；
减持	未来 6-12 个月的投资收益率落后市场基准指数 5%至 15%；
卖出	未来 6-12 个月的投资收益率落后市场基准指数 15%以上；
无评级	因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。
基准指数说明：A 股市场基准为沪深 300 指数；香港市场基准为恒生指数；美国市场基准为纳斯达克综合指数或标普 500 指数。	

分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，并对本报告的内容和观点负责。负责准备以及撰写本报告的所有研究人员在此保证，本研究报告中任何关于发行商或证券所发表的观点均如实反映研究人员的个人观点。研究人员获取报酬的评判因素包括研究的质量和准确性、客户反馈、竞争性因素以及光大证券股份有限公司的整体收益。所有研究人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中的具体的推荐意见或观点有直接或间接的联系。

法律主体声明

本报告由光大证券股份有限公司制作，光大证券股份有限公司具有中国证监会许可的证券投资咨询业务资格，负责本报告在中华人民共和国境内（仅为本报告目的，不包括港澳台）的分销。本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格编号已披露在报告首页。

中国光大证券国际有限公司是光大证券股份有限公司的关联机构。

特别声明

光大证券股份有限公司（以下简称“本公司”）成立于 1996 年，是中国证监会批准的首批三家创新试点证券公司之一，也是世界 500 强企业——中国光大集团股份公司的核心金融服务平台之一。根据中国证监会核发的经营证券期货业务许可，本公司的经营范围包括证券投资咨询业务。

本公司经营范围：证券经纪；证券投资咨询；与证券交易、证券投资活动有关的财务顾问；证券承销与保荐；证券自营；为期货公司提供中间介绍业务；证券投资基金代销；融资融券业务；中国证监会批准的其他业务。此外，本公司还通过全资或控股子公司开展资产管理、直接投资、期货、基金管理以及香港证券业务。

本报告由光大证券股份有限公司研究所（以下简称“光大证券研究所”）编写，以合法获得的我们相信为可靠、准确、完整的信息为基础，但不保证我们所获得的原始信息以及报告所载信息之准确性和完整性。光大证券研究所可能将不时补充、修订或更新有关信息，但不保证及时发布该等更新。

本报告中的资料、意见、预测均反映报告初次发布时光大证券研究所的判断，可能需随时进行调整且不予通知。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。客户应自主作出投资决策并自行承担投资风险。本报告中的信息或所表述的意见并未考虑到个别投资者的具体投资目的、财务状况以及特定需求。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。

不同时期，本公司可能会撰写并发布与本报告所载信息、建议及预测不一致的报告。本公司的销售人员、交易人员和其他专业人员可能会向客户提供与本报告中所载观点不同的口头或书面评论或交易策略。本公司的资产管理子公司、自营部门以及其他投资业务板块可能会独立做出与本报告的意见或建议不相一致的投资决策。本公司提醒投资者注意并理解投资证券及投资产品存在的风险，在做出投资决策前，建议投资者务必向专业人士咨询并谨慎抉择。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务。投资者应当充分考虑本公司及本公司附属机构就报告内容可能存在的利益冲突，勿将本报告作为投资决策的唯一信赖依据。

本报告根据中华人民共和国法律在中华人民共和国境内分发，仅向特定客户传送。本报告的版权仅归本公司所有，未经书面许可，任何机构和个人不得以任何形式、任何目的进行翻版、复制、转载、刊登、发表、篡改或引用。如因侵权行为给本公司造成任何直接或间接的损失，本公司保留追究一切法律责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

光大证券股份有限公司版权所有。保留一切权利。

光大证券研究所

上海

静安区新闻路 1508 号
静安国际广场 3 楼

北京

西城区复兴门外大街 6 号
光大大厦 17 层

深圳

福田区深南大道 6011 号
NEO 绿景纪元大厦 A 座 17 楼

光大证券股份有限公司关联机构

香港

中国光大证券国际有限公司
香港湾仔告士打道 108 号光大中心 33 楼