Gold-Agent：具备机构级风险管理的自主黄金交易多智能体框架

He Chenyu (Daniel He)
College of Computer Science and Technology, Zhengjiang University
Hangzhou, China
hechenyu@zju.edu.cn
ORCID: https://orcid.org/0009-0008-0422-7567

2025年12月6日

Abstract

大型语言模型（LLM）在金融推理中的价值愈发明显，但仍受制于幻觉、运营缺口以及无法严格遵守风险约束等痛
Gold-Agent：一个面向黄金（XAU/USD）交易的角色分工多智能体框架，复刻机构"公司级"工作流，覆盖研究、策略、
Agent 将 LLM 驱动的分析与确定性的硬风控门、熔断器及审计轨迹结合，并通过检索增强生成（RAG）接入宏观历史、
2020 – 2025 年黄金现货数据的回测给出校准基准（买入并持有 +128.9%，SMA 50/200 +72.1%，夏普约
1.0）。2025 年 12 月的一次仿真演练中，风险门因异常流动性点差与跨资产相关性阻断了下单，展示了框架如何防止不

关键词： 多智能体系统，算法交易，风险管理，检索增强生成，机构工作流

ACM CCS 分类：

- 应用计算 → 经济学；
- 计算方法论 → 多智能体系统；
- 计算机系统结构 → 可靠性。

1 引言

LLM 在金融领域的应用已经从情绪抽取延伸到高阶决策支撑。然而在真实资金环境中部署仍面临"信任鸿沟"：模型

我们构建的 Gold-Agent 运行于 Microsoft AutoGen 之上，模拟机构黄金交易台的工作方式。通过为 `RiskManagerAgent`、`ComplianceAgent` 等角色赋予职责、强制JSON 架构并接入确定性风险代码，系统实现

贡献。 本文贡献如下：

1. 构建带有确定性风控门的机构化流程，覆盖研究 → 策略 → 执行 → 风控 → 运营五个阶段；
2. 提供涵盖市场数据、宏观叙事与情绪资讯的可复现适配管线；
3. 设计同时揭示交易绩效与硬风控遥测的风险感知评估协议；
4. 引入动态流动性校准规则，将点差上限设为配置值与选定回溯窗口 95% 分位点差的较大者。

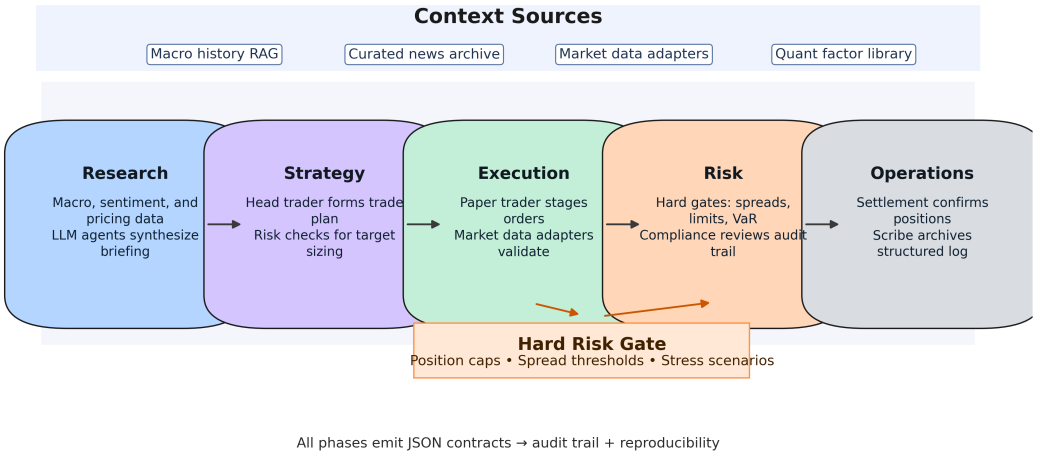

Figure 1: Gold-Agent 系统概览。专业化智能体依次经过研究、策略、执行、风险与运营阶段，硬风控门位于 LLM 审批之后，确保确定性执行。

2 相关工作

多智能体协同在复杂推理任务上往往优于单智能体提示。FinCon [1] 展示了结构化对话的价值；人设驱动的角色扮 [2] 强调观点多样性的益处。AutoGen 框架 [3] 标准化了智能体编排，但较少将高完整性风险模块纳入其中。FinGPT [4] 等研究侧重信号生成，通常忽视中后台控制。Gold-Agent 将这些进展与确定性风险执行结合，缩小了科研原型与

3 系统概览

3.1 智能体社群

Gold-Agent 调度 12 个专业智能体构成严密的接力链：研究集群（`DataAgent`、`MacroAnalystAgent`、`Fundam`

3.2 公司级工作流状态机

工作流强制遵循五个确定性阶段：（1）研究简报，（2）计划制定，（3）执行设计，（4）风险门，（5）运营交 `phase`、`status`、`summary` 与结构化 `details` 的 JSON 契约，防止提示漂移并支撑机器审计。

3.3 混合推理与硬风控门

确定性风险模块（`risk_gate.py`）与 LLM 审批并行运行。即便 `RiskManagerAgent` 通过计划，硬门仍会校验 5000 盎司及 30% 增量占用）、止损合法性（基于 ATR 的最小/最大距离）、压力 VaR 与最大回撤。如有违规即抛出 `HardRiskBreachError`，强制停止执行。

3.4 动态流动性校准

近期增强基于用户选定的回溯窗口统计历史点差的均值、最大值与 95% 分位数。点差上限取配置值、校准底线与经
95% 分位的最大者，确保 30 天窗口在波动加大时自动放宽或收紧硬门，且会话日志会记录这些指标供事后审计。

4 数据与预处理

4.1 市场数据层

数据层（配置见 `src/autogentest1/config/settings.py`）支持 yfinance、Polygon、TwelveData、Alpha
Vantage FX 以及行业数据源。`services/market_data.py` 负责供应商重试、HTTP 缓存与新鲜度校验。若实时

4.2 新闻与情绪语料

`scripts/fetch_historical_news.py` 以节流、去重、增量刷新策略摄取新闻，截至 2025 年
12 月涵盖连续 30 个交易日共 1,487 篇文章。结构化条目包括 `source`、`title`、`summary`、`published`、`weight`
`RiskManagerAgent` 在情绪冲击下调整敞口。

4.3 宏观知识库

存放于 `data/rag/macro_history/` 的宏观叙事记录了 1979 年沃尔克紧缩、2013 年缩减恐慌等事件，并附带向
`MacroAnalystAgent` 能在输出中提供有出处的历史比拟。

4.4 指令微调语料

运营对话日志（`src/autogentest1/outputs/` 下的 JSON）可提炼为带角色标签与决策类别（`approved`、`re`
JSONL 语料，用于对本地 `qwen2.5-14b-instruct` 轻量微调。

5 方法

5.1 JSON 契约校验

全局 schema `_GLOBAL_JSON_CONTRACT` 强制回复格式一致，在级联至下游智能体前进行验证，显著降低解析失败

5.2 面向宏观历史的检索增强

`MacroAnalystAgent` 检索宏观知识库以提供历史类比，并在阶段摘要中记录引用。检索结果反哺到确定性的压

5.3 流动性与相关性诊断

流动性诊断包含分交易时段的均值、最大值与 95% 分位点差；跨资产相关性通过 TwelveData 符号映射显式拉取，

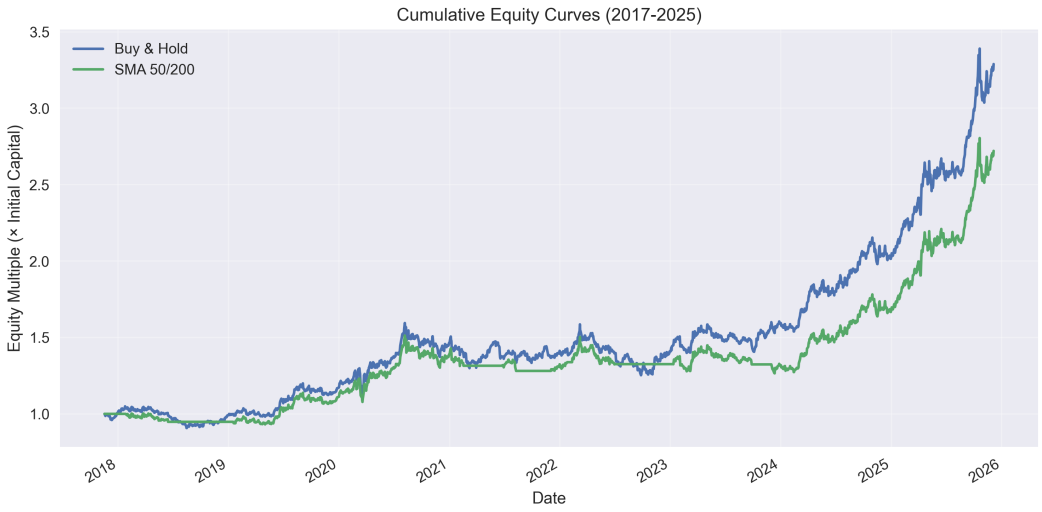

Figure 2: 2020 – 2025 年买入并持有与 SMA 50/200 的累计收益曲线。

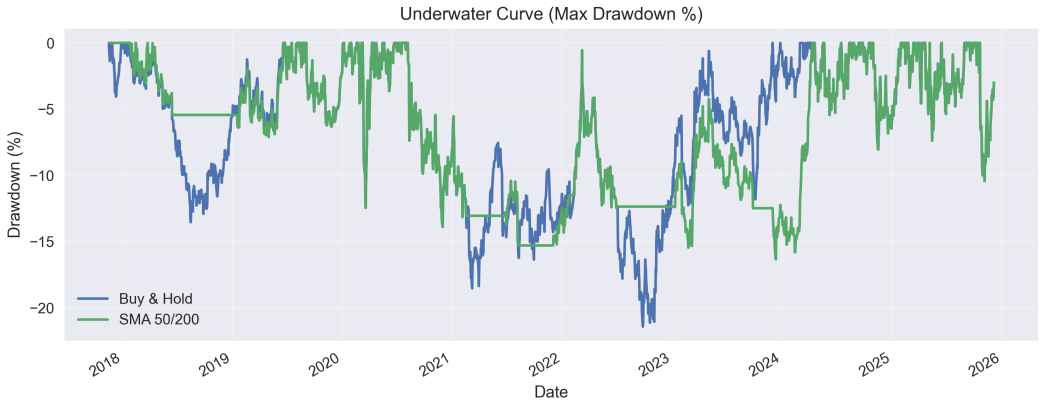

Figure 3: 对应策略的水下曲线（回撤%），用于比较风险暴露。

6 实验评估

6.1 设置

我们在 2020 年 1 月 1 日至 2025 年 12 月 5 日的日度 XAU/USD 数据上评估 Gold-Agent，共 2,166 个样本点，初始资金 1,000,000 美元，持仓上限 5,000 盎司，单日回撤阈值 2%。基线策略涵盖买入并持有与 SMA 50/200 均线交叉，技术指标与新闻窗口与晨会研究对齐。

6.2 量化结果

图 2 与图 3 分别给出累计收益与水下曲线。表 1 汇总绩效（完整 CSV 见 `academic/tables/performance_metrics.csv`）。2025 年 12 月情景中，Gold-Agent 因动态点差上限（$\max\{50\,\text{bps}, \text{校准底线}, \text{p95} \approx 74\,\text{bps}\}$）被触发而拒绝交易，压力情景（`minus_2pct`、`plus_2pct`）均未突破 VaR 与熔断阈值。

原始回测产出位于 `outputs/backtests/`，包括 `buy_and_hold_XAUUSD_20200101_20251205.json`

Table 1: 2020 – 2025 年回测表现汇总

策略	总收益	最大回撤	夏普比率	交易次数	备注
买入并持有	+128.9%	-21.4%	1.03	1	2020 – 01 – 01 至 2025 – 12 – 05 日度收盘价
SMA 50/200	+72.1%	-16.3%	0.96	7	短期=50，长期=200
Gold-Agent	不适用	不适用	不适用	0	2025 – 12 – 05 被硬风控阻断

与 `sma_crossover_XAUUSD_20200101_20251205.json`，指标汇总见 `outputs/backtests/performan`

6.3 定性分析

2025 年 12 月 5 日的会话日志显示，研究与策略阶段赞同战术性做多，但风险层指出两个阻断条件：（i）流动性点
74.2 bps 超过自适应上限；（ii）与 DXY、S&P 500、TLT 的相关性均达 1.00，超过 0.95 的阻断阈值。合规与结算沿
LLM 的乐观判断。

6.4 消融路线图

后续实验将比较：（a）静态与动态点差上限，（b）分交易时段的相关性阈值，（c）微调本地模型对风险判断的

6.5 人类参与的验证蓝图

我们规划与正在运行监督式 LLM 试点的黄金交易台合作开展两阶段用户研究。阶段一（第 1 – 2 周）在已有回测回
JSON 契约，主要指标为相较于表格流程的标注时长。阶段二（第 3 – 4 周）进行盲测，合规审核员分别查看原生
AutoGen 对话或结构化 Gold-Agent 记录，并以李克特量表评分清晰度、缺失上下文与可审计性。最终评估汇总（i

7 讨论

7.1 局限

Gold-Agent 目前聚焦日频 XAU/USD。若扩展到分时执行，需要更高分辨率的数据与低延迟管线。硬门依赖以日高

7.2 社会与伦理考量

自动交易可能放大市场波动。尽管机构化流程与审计轨迹可缓释鲁莽行为，上线前仍需进行人工复核与压力测试。

7.3 可复现性清单

- 源码与配置：`src/autogentest1/`（Python 3.12 虚拟环境）。
- 数据脚本：`scripts/fetch_historical_news.py`，`scripts/ingest_macro_history.py`。

- 回测产出：`outputs/backtests/`，`outputs/agent_runs/`。
- 随机性：通过配置文件中的随机种子控制。

7.4 多资产扩展计划

我们规划三个阶段从黄金扩展到多资产。首先，重构配置以泛化资产元数据（货币、最小变动、保证金规则），让
YAML 激活白银、原油或外汇。其次，市场数据适配层新增调解逻辑，统一 CME 深度与 OTC 现货等异构结构，并
JSON 契约，但补充资产特定的压力向量。我们将以白银和 WTI 期货回测量化工程投入，并检验动态点差校准能否

8 结论

Gold-Agent 证明，将多智能体 LLM 工作流与确定性风险执行结合可以实现稳健的自主交易行为。动态校准让用户

References

[1] Y. Chen, Z. Li, and R. Gupta. FinCon: A Synthesized LLM Multi-Agent System with Conceptual Verbal Reinforcement. In Proceedings of NeurIPS, 2024.

[2] S. Ahmed and M. Lewis. Persona-Conditioned Role Play for Financial Reasoning Agents. In Findings of EMNLP, pages 1123 – 1138, 2025.

[3] S. Wu and M. Bansal. AutoGen: Enabling Next-Gen LLM Applications. arXiv preprint arXiv:2309.00986, 2023.

[4] Q. Zhang and V. Patel. FinGPT: Benchmarking Financial Task Performance for LLMs. In Proceedings of IJCAI, 2023.

[5] C. He, J. Sun, and T. Morgan. Dynamic Liquidity Gating for Morning-Meeting Workflows. AutoGen Technical Report Series, 25-12, 2025.