大语言模型时代的金融智能体

大语言模型的能力评测、结构化建模和交易算法设计

刘晨

香港城市大学　计算机科学系



2026 年 11 月 15 日
人工智能金融创新论坛，上海

金融数据的爆炸式增长与其复杂性

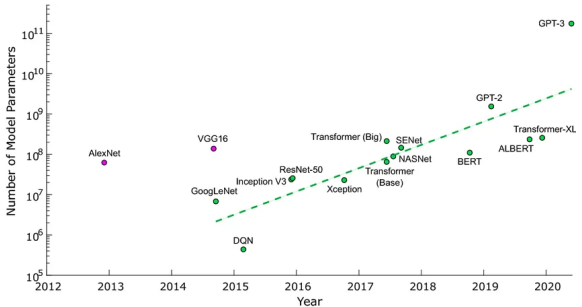
近年来，金融数据规模庞大并持续增长，呈现 “量大异构” 的特点。

- **高频市场数据**：实时（tick 级）交易数据每天生成 PB 级信息。
- **监管与宏观数据**：这些数据来自中央银行或报告机构。它还包括全球经济和金融健康状况的宏观特征。
- **媒体数据**：这些数据量大并较能反映市场主体的态度和预期，但是主观性较强并且存在自相矛盾的情况。
- **多模态另类数据**：机构现在纳入大量非传统数据集，例如用于供应链跟踪的卫星图像、用于零售客流量的地理位置数据，以及来自物流的物联网传感器数据，以获得信息优势。

大数据驱动的人工智能

过去十年见证了人工智能的巨大发展，尤其是深度学习和大语言模型的发展。与十年前的人工智能相比，当今的人工智能系统，即多模态大语言模型（LLMs），具有以下特性：

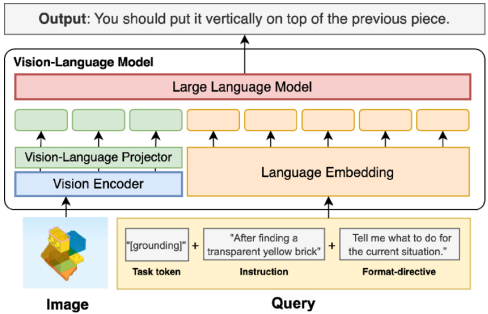
- **强大的表达能力**：前所未有的参数规模使这些模型能够有效逼近复杂的高维分布，并编码海量世界知识。



大数据驱动的人工智能

过去十年见证了人工智能的巨大发展，尤其是深度学习和大语言模型的发展。与十年前的人工智能相比，当今的人工智能系统，即多模态大语言模型（LLMs），具有以下特性：

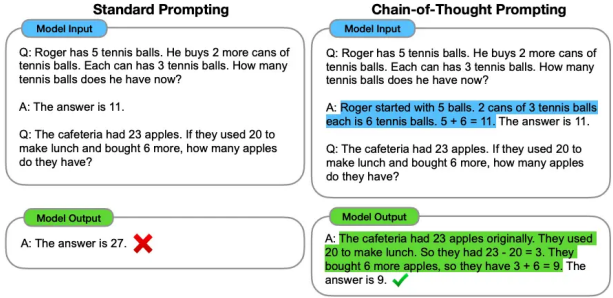
- **多模态**：通过利用映射到共享潜在空间的专用编码器，系统可以处理、对齐并合成多模态数据。



大数据驱动的人工智能

过去十年见证了人工智能的巨大发展，尤其是深度学习和大语言模型的发展。与十年前的人工智能相比，当今的人工智能系统，即多模态大语言模型（LLMs），具有以下特性：

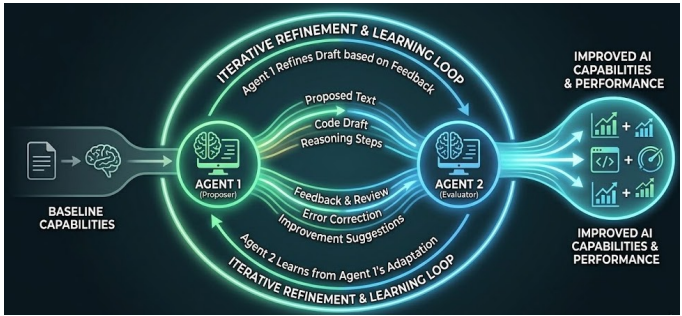
- **推理**：除了基本模式识别之外，这些模型还展现出多步逻辑推演能力。通过思维链（Chain-of-Thought, CoT）等机制，它们可以分解复杂问题并分析相互竞争的信号。



大数据驱动的人工智能

过去十年见证了人工智能的巨大发展，尤其是深度学习和大语言模型的发展。与十年前的人工智能相比，当今的人工智能系统，即多模态大语言模型（LLMs），具有以下特性：

- **自我改进**：由 LLM 驱动的 AI 智能体可以相互交互以自我改进。



人工智能可以赋能金融科技

近年来人工智能的发展，尤其是大模型，具备独特能力来推动金融科技向前发展。

- **海量复杂数据** ⇐ 大模型的扩展上下文窗口能力。
- **多样数据模态** ⇐ 大模型的原生多模态处理架构。
- **异构数据源** ⇐ 在编码广泛世界知识方面具有前所未有的能力。
- **可解释性需求** ⇐ 透明、逐步推理能力。
- **噪声大且波动强的信号** ⇐ 鲁棒、交叉验证的多智能体系统。

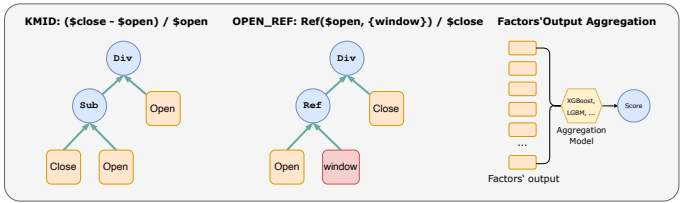
近两年的主要工作

1. **股票：Alpha 因子挖掘**
   - AlphaBench：模型能否生成、判断和搜索因子?
   - Evolutionary Factor Search（EFS）：用回测反馈指导因子进化。
2. **期权：交易与建模**
   - LiveOption：智能体能否构造并执行合适的期权策略?
   - AddMart：用稀疏报价学习符合无套利条件的 IV 曲面。

Alpha 因子：把市场历史变成选股信号

一个因子是一条**可计算的公式**，把价格、成交量等历史信息转换成股票得分。

- 直观例子：近期涨幅是否预示趋势延续，或短期反转?
- 按得分给股票排序，再结合风险约束构建投资组合。



关键检验：排序能否在未来数据上成立?

AlphaBench: Benchmarking Large Language Models in Formulaic Alpha Factor Mining，ICLR 2026.

AlphaBench：拆解因子研究中的模型能力

任务	模型做什么	检验什么
因子生成	把自然语言想法写成公式	能执行，且符合研究意图
因子评估	不做回测，直接判断因子好坏	评分和排序是否可靠
因子搜索	根据回测反馈迭代修改公式	改进幅度与搜索成本

统一表达式语言、执行环境和回测流程，支持可比的模型与搜索评测。

把生成、判断与迭代改进三项能力分别测清楚。

AlphaBench: Benchmarking Large Language Models in Formulaic Alpha Factor Mining，ICLR 2026.

AlphaBench：生成能力强于质量判断能力

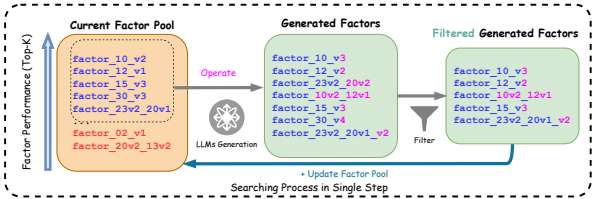
- **能写出合法公式，但理解研究意图仍有差距。**
  例如 GPT-5 的生成可靠性为 1.00，语义准确性为 0.56。
- **仅凭公式判断投资价值，表现普遍较弱。**
  零样本信号识别与两两选择的准确率通常接近随机水平。
- **在推理中引入更长的思维链，未带来一致的改善。**
  搜索 Alpha 因子的算法依然十分重要。

适合让 LLM 提出候选，再由数据反馈检验候选。

AlphaBench: Benchmarking Large Language Models in Formulaic Alpha Factor Mining，ICLR 2026.

Evolutionary Factor Search (EFS)：用回测反馈指导因子进化

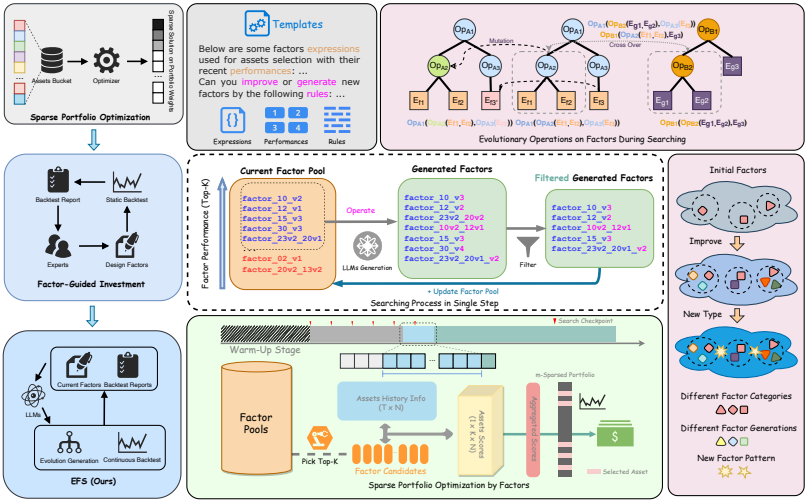
核心思路：维护一个因子种群，让 LLM 结合已有公式和表现，提出变异与组合。



- **提出候选**：修改表达式结构，并给出金融直觉。
- **回测筛选**：保留表现较好的候选，持续更新因子池。
- **形成组合**：聚合因子得分，选择少量股票进行配置。

EFS: Evolutionary Factor Searching for Sparse Portfolio Optimization Using Large Language Models.

Evolutionary Factor Search (EFS): 用回测反馈指导因子进化



EFS: Evolutionary Factor Searching for Sparse Portfolio Optimization Using Large Language Models.

Evolutionary Factor Search (EFS)：大模型的优势

用大模型充当进化算法主循环的引擎有以下独特的优势：

- **广泛上下文**：大模型编码了广泛领域的文献。
- **新颖性搜索**：大模型的创造力能够识别空白并提出新颖策略。
- **代码生成**：将伪代码转换为可执行的 Python 脚本。
- **工具交互**：大模型智能体可以与工具 / 软件交互以运行代码。
- **自我改进**：经验结果可以作为下一轮迭代的提示输入大模型，以实现自我改进。

EFS: Evolutionary Factor Searching for Sparse Portfolio Optimization Using Large Language Models.

Evolutionary Factor Search (EFS)：回测显示潜力，部署仍需验证

Table: 美国（US50）、香港（HSI45）和中国内地（沪深 300）市场上的累积回报 (CW↑), 夏普比率 (SR↑), 和最大回撤 (MDD↓) 对比。美国市场为 2019 到 2024 年，香港和中国内地市场为 2022 到 2025 年

		US50			HSI45			CSI300		
Group	Method	CW↑	SR↑	MDD↓	CW↑	SR↑	MDD↓	CW↑	SR↑	MDD↓
Baseline	1/N	4.562	0.072	0.344	1.333	0.029	0.409	1.087	0.014	0.214
	Min-cVaR	1.779	0.038	0.314	1.628	0.063	0.244	0.992	0.003	0.286
	Max-Sharpe	4.495	0.061	0.461	1.428	0.043	0.300	1.008	0.007	0.333
m=10	LGBM	4.182	0.063	0.332	1.611	0.038	0.367	2.334	0.072	0.225
	XGBoost	6.313	0.077	0.328	1.581	0.035	0.440	1.420	0.032	0.345
	mSSRM-PGA	5.121	0.059	0.569	0.766	-0.003	0.547	0.881	0.002	0.399
	ASMCVaR	10.259	0.073	0.582	2.481	0.052	0.453	1.453	0.030	0.462
	EFS-DeepSeek	**25.101**	0.132	0.288	**3.463**	0.080	0.385	3.437	0.079	0.327
	EFS-GPT	22.905	0.130	0.260	2.789	0.067	0.292	**4.962**	0.098	0.301
m=15	LGBM	3.899	0.062	0.328	1.588	0.037	0.387	1.812	0.055	0.250
	XGBoost	5.607	0.076	0.319	1.586	0.036	0.420	1.348	0.029	0.344
	mSSRM-PGA	4.976	0.062	0.477	0.766	-0.003	0.547	0.787	-0.010	0.384
	ASMCVaR	11.124	0.074	0.566	**2.647**	0.054	0.434	1.658	0.035	0.424
	EFS-DeepSeek	13.978	0.114	0.298	2.364	0.061	0.406	2.510	0.067	0.298
	EFS-GPT	**14.707**	0.117	0.278	2.277	0.058	0.307	**3.218**	0.082	0.246

EFS: Evolutionary Factor Searching for Sparse Portfolio Optimization Using Large Language Models.

期权交易：同时决定方向、期限与收益结构

与股票相比，期权的自动化交易需要克服**非线性收益**、**不可观测的波动率**、**时间维度**、**更复杂的对冲机制**等挑战。

一个看涨判断，还不足以确定交易。

- 买哪一种期权，选什么行权价?
- 多久到期，买多少，是否组合多腿?
- 如何满足资金与保证金约束?

即使方向看对，也可能亏损。

- 标的涨幅不足以覆盖权利金。
- 时间流逝使期权价值衰减。
- 隐含波动率下降压低期权价格。

两项互补能力：LiveOption 检验**交易决策**，AddMart 支持**定价与风险建模**。

LiveOption：期权智能体的三类任务评测

场景	交易目的	核心挑战
组合叠加策略	为持仓对冲或收取权利金	保护成本与收益的权衡
财报事件交易	围绕财报交易方向或波动	事件冲击与波动率变化
日内 0DTE	交易当日到期的期权	快速衰减与风险控制

- 分析智能体形成交易意图，执行智能体选择合约与数量。
- 环境检查订单、资金与执行规则，并记录完整决策过程。

LiveOption: Evaluating LLM Agents in Structured Option Trading with Nonlinear Payoffs，NeurIPS 2026.

LiveOption：收益、风险与行为需要一起看

- **收益结果：当前智能体尚未表现出稳定的交易优势。**
  财报与 0DTE 两类任务中，所有模型的收益中位数均为负。
- **风险结果：部分对冲降低了回撤，但同时付出收益成本。**
  应结合任务目标评价保护效果，不能只按最终收益排名。
- **行为诊断：检查合约、仓位与交易成本。**
  0DTE 少数模型取得的正平均收益，对更真实的执行成本十分敏感。

有效的评测既要解释赚了多少，也要解释承担了什么风险。

LiveOption: Evaluating LLM Agents in Structured Option Trading with Nonlinear Payoffs，NeurIPS 2026.

AddMart：用少量报价重建 IV 曲面

- **隐含波动率（IV）**：把市场期权价格换算成统一的波动率口径。
- **曲面的两个坐标**：行权价相对标的的位置，以及剩余到期时间。
- **重建任务**：市场只给出离散报价，定价和风控需要连续曲面。

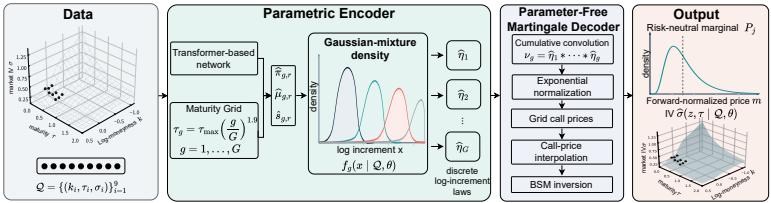
$$\mathrm{IV}(k,\tau) \qquad k\text{：相对行权价} \qquad \tau\text{：剩余期限}$$

曲面既要贴近报价，也要满足不同合约之间的无套利关系。

AddMart: Arbitrage-Free Implied Volatility Surface Reconstruction via Additive Martingales.

AddMart：把无套利条件写进模型结构

核心思路：学习合理的价格分布与鞅（Martingale）过程，再恢复期权价格和 IV 曲面。



- **少量输入**：主实验用 9 个报价，重建其余合约的 IV。
- **结构约束**：保证风险中性分布有效，并协调不同行权价与期限。
- **实验结果**： 10 个数据集上精度有竞争力，未见实质性静态套利违约。

AddMart: Arbitrage-Free Implied Volatility Surface Reconstruction via Additive Martingales.

总结与展望

总结

- 大语言模型给处理大量异构的金融数据提供了便利。
- 直接使用通用大语言模型进行量化交易的结果并不理想，目前我们仍然需要设计专门的算法，合理且巧妙地利用大模型，结合市场与交易对象的特性，构建自动化交易系统。

展望

- 更加高效（时间延迟更小、消耗 Token 数更少）的交易算法。
- 构建多智能体评测框架，用于模拟市场上多个主体之间的交易。探究一个大量使用基于大模型交易算法的市场的特性。

谢谢！

致谢： 张元，罗浩琛，陈建东，罗晓龙。

参考文献：

- AlphaBench: Benchmarking Large Language Models in Formulaic Alpha Factor Mining.
- EFS: Evolutionary Factor Searching for Sparse Portfolio Optimization Using Large Language Models.
- LiveOption: Evaluating LLM Agents in Structured Option Trading with Nonlinear Payoffs.
- AddMart: Arbitrage-Free Implied Volatility Surface Reconstruction via Additive Martingales.