

量化交易新范式：事件驱动策略与另类数据的深度融合与前沿应用分析报告

1. 宏观背景：量化投资的代际跃迁

全球金融市场正处于一场深刻的结构性变革之中。随着高频交易(HFT)竞争的白热化和传统因子(如动量、价值)的拥挤度不断攀升，超额收益(Alpha)的获取变得愈发艰难。在此背景下，量化投资已从单纯依赖价格和财务报表的“统计套利1.0”时代，全面迈向利用非结构化数据和复杂事件逻辑的“另类量化2.0”时代。

事件驱动(Event-Driven)策略，这一曾经由主观基本面基金经理主导的领域，正迅速被量化机构通过自然语言处理(NLP)、计算机视觉(CV)和知识图谱等前沿技术攻占。与此同时，另类数据(Alternative Data)的爆发式增长为这些策略提供了前所未有的信息燃料。本报告将深入剖析这一融合趋势，从策略原理、数据技术、基础设施挑战到合规边界，进行详尽的系统性梳理与分析。

2. 事件驱动策略的量化重构

事件驱动投资的核心在于利用企业重大事件引发的市场定价效率低下。这些事件通常具有明确的时间催化剂(Hard Catalyst)或较软的预期催化剂(Soft Catalyst)，导致资产价格在短时间内脱离其基本面价值或产生结构性重估。

2.1 系统化并购套利(Systematic Merger Arbitrage)

并购套利(Merger Arbitrage)是事件驱动策略中最古老且最基础的形式，其本质是捕捉目标公司股价与收购要约价格之间的价差(Spread)。在量化视角下，这不再是单一的交易判断，而是一个复杂的风险溢价模型。

2.1.1 策略逻辑与收益来源

当并购消息公布时，目标公司的股价通常会大幅上涨，但往往会略低于收购报价。这一折价反映了交易失败的风险(Deal Break Risk)以及资金的时间价值。量化并购套利策略通过做多目标公司股票，同时(如果是换股交易)做空收购方股票来锁定这一价差¹。

从风险因子的角度来看，并购套利策略表现出一种类似“看跌期权”的收益特征：在大多数情况下赚取稳定的价差收益(Carry)，但在交易失败或市场崩盘时面临显著的下行风险。研究表明，HFRI事件驱动指数与HFRI并购套利指数在2020年至2025年间表现出高达0.9的相关性，且这种策略在市场压力时期往往表现出对权益市场(Equity Beta)的隐性暴露²。

2.1.2 量化建模的演进

现代量化基金不再依赖人工判断反垄断风险，而是构建多因子模型来预测交易完成的概率(

Implied Probability of Completion)。核心变量包括：

- 溢价率(Premium)分析:过高的溢价可能暗示收购方出价过高,增加了股东否决的风险;过低的溢价则可能引发竞争性报价。
- 交易结构特征:全现金、全股票或混合支付方式对套利价差的动态影响显著不同。现金收购通常受利率环境影响较大,而换股收购则受收购方股价波动影响。
- 监管风险量化:利用NLP技术实时扫描反垄断机构(如FTC、欧盟委员会)的公告和相关新闻,量化监管干预的可能性。例如,对于涉及科技巨头的并购案,模型会自动提高所需的风险溢价阈值。
- 宏观环境相关性:并购活动具有顺周期性。在经济强劲时期,并购交易量大,成功率高;而在衰退期,不仅交易量萎缩,已宣布的交易也更容易因融资困难而告吹¹。

2.2 特殊情况投资(Special Situations)

除了标准的并购套利,量化策略正深入挖掘更复杂的公司行为。

2.2.1 分拆上市(Spinoffs)

分拆是指母公司将子公司剥离为独立上市实体。这类事件通常会引发显著的非基本面抛售压力。例如,如果母公司是标普500指数成分股,而被分拆的子公司市值较小未被纳入指数,指数基金将被迫在分拆完成后抛售子公司股票。量化模型通过预测这种机械性的资金流动(Flow-driven pressure),在抛售压力最大时买入,待价格回归基本面后卖出³。

2.2.2 困境证券与破产重组(Distressed & Bankruptcy)

困境投资传统上是极度依赖法律专家的领域。然而,随着法律文档的数字化,量化基金开始尝试介入。

- 文本分析:通过NLP分析破产法庭的通过文件(Court Dockets),提取关键的重组条款、债权人情绪和法官倾向,从而预测清算回收率(Recovery Rate)。
- 反周期特性:与并购套利不同,困境投资在经济衰退期表现最佳,因为此时违约率上升,资产错杀机会增多¹。

2.3 盈余公告与预期管理(Earnings Surprises)

盈余公告后漂移(Post-Earnings Announcement Drift, PEAD)是市场上最持久的异象之一。现代量化策略的重点已从单纯的“反应”(Reaction)转向“预测”(Prediction)。

2.3.1 超越共识预期(Beyond Consensus)

传统的量化模型使用IBES等数据库的分析师一致预期(Consensus Estimate)来计算盈余惊喜(SUE)。然而,这种数据往往存在滞后性。量化机构现在利用“智能预期”(Smart Estimates),即根据分析师的历史准确度、发布时间新旧进行加权,构建更准确的预期基准⁴。

2.3.2 另类数据的预测力

更进一步,量化基金通过接入信用卡消费数据、App下载量、甚至供应链进出口数据,在财报发布

前几周构建独立的营收预测模型。如果量化模型的预测值显著偏离市场一致预期，这便构成了一个强有力的交易信号。实证研究显示，利用这种“信息差”构建的策略，在财报窗口期能获得显著的超额收益³。

3. 另类数据 (Alternative Data) 的经济学与生态系统

另类数据是指那些非来源于公司财报、交易所行情或政府统计局的非传统数据。它已成为量化基金寻求差异化竞争优势的核心战场。

3.1 市场规模与预算结构

另类数据市场正经历爆发式增长。据预测，到2025年，全球另类数据市场规模将达到187.4亿美元，并预计以超过63%的复合年增长率在2030年达到1358亿美元⁶。这种增长背后是买方机构预算的急剧扩张。

- 高昂的成本:大型多策略对冲基金平均每年在另类数据上的支出约为500万美元，通常订阅超过40家不同的数据供应商。即使是较小的基金，平均支出也达到160万美元。单个高质量数据集的年订阅费平均约为8万美元⁷。
- 护城河效应:这种高昂的资金门槛不仅筛选了参与者，也加剧了头部效应。拥有更多数据预算的基金能够挖掘更独特 (Unique) 的信号，从而获得更持久的Alpha。

3.2 数据的生命周期: 从Alpha到Beta

另类数据不仅昂贵，而且存在显著的“阿尔法衰减”(Alpha Decay)现象。

- 独特性与拥挤度: 当一个数据集 (如早期的信用卡数据) 仅被少数基金使用时，它能提供巨大的超额收益。但随着数据供应商客户群的扩大，该信号迅速被市场定价，转化为一种“另类Beta”。研究表明，另类数据的Alpha优势通常能维持约4年⁸。
- 持续挖掘的压力: 为了对抗衰减，量化团队必须不断寻找新的、未被开发的数据源 (如特定行业的物联网传感器数据、利基市场的网络流量等)，或者通过更先进的处理技术 (如深度学习) 从现有数据中提取更深层的特征⁹。

3.3 数据分类学 (Taxonomy)

另类数据种类繁多，通常可分为以下几类：

类别	来源示例	量化应用场景	处理难度
个人产生数据	社交媒体 (Twitter, Reddit)、产品评论、论坛发帖	情绪分析、散户逼空预测、品牌健康度监测	高 (非结构化文本、噪音大)

商业流程数据	信用卡交易、电子收据、供应链物流单据	营收预测、库存周转分析、宏观消费趋势	中 (结构化但需脱敏清洗)
传感器数据	卫星图像、手机GPS定位、物联网 (IoT) 设备	零售客流计数、原油库存监测、农作物产量预估	极高 (海量图像/信号处理)
企业废气数据	政府合同、专利申请、招聘信息、法律文件	创新能力评估、扩张计划预测、法律风险预警	中 (需专业领域知识)

4. 自然语言处理(NLP)与文本挖掘的深度应用

在所有另类数据中，文本数据(Text Data)体量最大、蕴含信息最丰富。从早期的词频统计到如今的大语言模型(LLM)，NLP技术彻底改变了量化投资者理解市场的方式。

4.1 技术演进：从FinBERT到生成式Agents

4.1.1 专用小模型 vs. 通用大模型

- **BERT与FinBERT**: 在2019-2023年间，基于Transformer架构的BERT模型是主流。特别是专门针对金融语料库(如财报、财经新闻)微调的FinBERT，在金融情绪分类任务上显著优于通用模型。实证数据显示，FinBERT在Twitter金融新闻情绪分类上的准确率可达88.4%¹⁰。其优势在于推理成本低、速度快，适合处理海量实时新闻流。
- **LLM的崛起**: 进入2024-2025年，以GPT-4、DeepSeek-R1、Llama-3为代表的大语言模型展现了惊人的零样本(Zero-shot)推理能力。研究表明，经过精心设计的提示工程(Prompt Engineering)，ChatGPT在某些外汇情绪判断任务上比FinBERT高出35%的性能¹¹。LLM不仅能判断情绪的正负，还能识别反讽、进行多步逻辑推理(Chain-of-Thought)以及生成摘要¹²。

4.1.2 混合架构

目前的最佳实践往往是混合架构：利用轻量级的FinBERT或DistilRoBERTa处理高频、低延迟的新闻流筛选，而将复杂、低频的深度分析(如长篇财报电话会议纪要)交给LLM处理，以平衡成本与精度¹⁰。

4.2 深度应用场景解析

4.2.1 央行沟通与宏观情绪

美联储(Fed)及全球央行的会议纪要、官员讲话是影响资产价格的关键宏观变量。NLP模型被用来量化这些文本中的“鹰派”(Hawkish)或“鸽派”(Dovish)程度。通过对比前后两次声明的语调变化(

Tone Shift), 模型可以预测利率路径的微调。DeepSeek-R1等模型在无需微调的情况下, 即可精确提取目标级(Target-level)情绪, 帮助宏观对冲基金调整债券和外汇仓位¹³。

4.2.2 财报电话会议(Earnings Calls)的高维分析

除了文本内容, 现代NLP还结合了音频分析。

- 语言复杂度: 管理层如果使用复杂的句式、模糊的词汇或过多的“犹豫词”(如uh, um), 往往暗示信心不足或试图掩盖负面信息。
- 问答环节(Q&A): 这部分通常比照本宣科的陈述环节包含更多Alpha。量化模型分析分析师提问的尖锐程度以及管理层回答的直接程度。研究发现, 电话会议中的语调特征与未来股价表现呈现显著正相关¹⁴。

4.2.3 员工情绪与并购预测(M&A Prediction)

这是一个典型的“非显性”信号应用。通过分析Glassdoor和LinkedIn上的现任及前任员工评论, 量化模型可以构建“员工满意度”和“企业文化”因子。

- 预测逻辑: 员工情绪的急剧下降可能预示着管理混乱或业绩压力, 这有时反而会增加公司“寻求出售”的可能性。实证研究表明, 结合新闻情绪和员工评论的机器学习模型, 在预测潜在并购目标(M&A Targets)方面, 比单纯依靠财务指标的传统统计模型更为准确。特别是负面情绪的积累, 在特定情境下对并购概率有正向预测力¹⁵。

4.2.4 社交媒体与散户逼空

WallStreetBets事件后, 监测Reddit、X(Twitter)等平台的散户情绪成为标配。NLP模型需要识别非标准的金融术语(如"Diamond Hands", "YOLO")以及表情符号(火箭、月亮)。通过量化特定股票的提及量(Mention Volume)和情绪极性, 模型可以预警流动性驱动的逼空风险, 或捕捉动量交易机会¹⁰。

5. 计算机视觉(CV)与地理空间智能

“从太空看地球”赋予了量化投资者上帝视角。计算机视觉技术将海量的卫星图像转化为结构化的时间序列数据, 成为大宗商品和零售行业的重要领先指标。

5.1 零售客流: 数车策略(Car Counting)

这是卫星数据最经典的金融应用之一。

- 执行机制: 通过高分辨率卫星图像, 算法自动识别并计数大型零售商(如Walmart, Home Depot, Lowe's)露天停车场的车辆数量。
- Alpha来源: 客流量是线下零售收入的直接代理变量。加州大学伯克利分校的研究发现, 基于卫星图像构建的交易策略, 在零售商季度盈余公告日前后三天的窗口期内, 能产生4%-5%的超额收益。年化回报率惊人⁵。
- 局限与进化: 随着这一策略的普及, 其Alpha正在衰减。现在的模型需要更精细化, 例如区分

员工车辆与顾客车辆，或者结合手机GPS定位数据(Foot Traffic)进行交叉验证，以提高预测精度¹⁸。

5.2 大宗商品：从油罐阴影到作物光谱

5.2.1 原油库存监测

传统的官方库存数据(如EIA报告)存在滞后性，且覆盖范围有限。Orbital Insight等公司利用计算机视觉技术监测全球超过20,000个储油罐。

- **技术原理**：绝大多数大型储油罐采用浮顶设计(Floating Roof)，顶盖随油位升降。卫星图像通过分析罐壁投射在顶盖上的阴影形状和角度(Shadow Analysis)，利用三角函数精确计算出油罐的存储量。
- **市场影响**：这种技术揭示了官方数据统计之外的隐形库存(特别是新兴市场如中国的商业库存)，对原油期货定价产生重大影响²⁰。

5.2.2 农业产量预测

这直接关系到农产品期货(玉米、大豆、小麦)的交易。

- **多光谱分析**：利用卫星的近红外波段计算归一化植被指数(NDVI)，监测作物的生长健康状况。
- **产量模型**：结合历史单产数据、土壤湿度微波遥感数据及天气预报，构建回归模型。例如，在斯里兰卡的稻米产量预测案例中，随机森林模型结合卫星数据实现了极高的精度，误差率控制在20%以内，且能比官方统计提前数周获得数据²¹。
- **实时与覆盖**：相比于传统的地面抽样调查，卫星遥感提供了全覆盖、无死角的监控能力，特别是在地缘政治动荡或交通不便的区域²¹。

6. 交易数据与消费者行为洞察

交易数据(Transaction Data)来源于信用卡、借记卡流水及电子收据，被公认为最接近“真相”的数据源。它允许投资者在公司发布财报前，就精确构建出其损益表(P&L)。

6.1 数据清洗与偏差修正

尽管价值巨大，原始交易数据却是极其“脏”且有偏的。

- **面板偏差(Panel Bias)**：数据提供商的样本通常来自特定的银行或记账App，其用户群体可能偏向年轻化、城市化或特定收入阶层。直接使用原始数据会导致严重误判。量化团队必须使用人口统计学方法(Demographic Scaling)及加权算法，将样本数据映射到全社会消费行为²³。
- **实体解析(Entity Resolution)**：信用卡账单上的商户描述符(Descriptor)通常是混乱的缩写(如"AMZN MKTP WA", "SBUX 1234")。算法需要将成千上万种变体准确映射到对应的上市公司(Ticker)及其具体子品牌。
- **清洗流程**：必须剔除退款>Returns)、拒付交易以及非消费性的转账。此外，如何处理借记卡

与信用卡的消费习惯差异也是关键²⁴。

6.2 策略应用：即时预测(Nowcasting)

利用清洗后的交易数据，量化基金进行“即时预测”。

- **营收惊奇(Revenue Surprise)**：在财报季，如果交易数据模型显示的季度营收增长为15%，而华尔街一致预期仅为5%，这便是一个强烈的做多信号。
- **宏观洞察**：除了个股，聚合的交易数据也是监测宏观经济(如零售销售总额、服务业复苏情况)的高频指标，帮助宏观对冲基金调整大类资产配置²⁶。

7. 知识图谱与供应链金融的复杂网络

世界是普遍联系的。传统的量化模型往往孤立地看待每个资产，而知识图谱(Knowledge Graphs, KG)和图神经网络(Graph Neural Networks, GNN)则通过显式建模实体间的关系，捕捉风险和收益在复杂网络中的传导。

7.1 供应链可视化与风险传导

现代供应链是高度全球化且脆弱的。一个底层的元器件供应商发生火灾，可能会导致数千公里外的汽车制造商停产。

- **图谱构建**：利用NLP技术从财报(10-K文件)、新闻报道和海关提单数据中抽取实体关系(如“供应商-客户”、“母公司-子公司”、“竞争对手”)，构建庞大的企业关系图谱²⁷。
- **图神经网络(GNN)应用**：GNN(如Graph Convolutional Networks, GCN)能够聚合邻居节点的信息来更新中心节点的特征。在供应链金融中，GNN被用于：
 1. **需求传导预测**：通过上游供应商的订单波动，利用图结构提前预测下游制造商的营收变化。研究表明，在股票预测任务中，考虑图结构特征(Edge-level features)的模型显著优于孤立的传统机器学习模型²⁹。
 2. **连锁反应模拟(Ripple Effect)**：模拟特定节点(如港口、工厂)失效后，风险如何沿供应链网络扩散，从而识别那些看似稳健但实则脆弱的公司³²。

7.2 欺诈检测与隐藏关系挖掘

在供应链金融中，欺诈往往涉及复杂的循环交易或虚假关联方。知识图谱能够通过图算法(如连通分量分析、PageRank)识别异常的资金闭环。例如，最新的双注意力时空图神经网络(DAST-GNN)通过同时建模时间和空间特征，在供应链金融欺诈检测任务中实现了93.64%的AUC值，比传统机器学习方法高出10%以上³⁴。

8. 量化基础设施与回测陷阱

引入另类数据对量化基金的技术架构提出了严峻挑战。

8.1 回测中的致命伤：前视偏差 (Look-ahead Bias)

前视偏差是指在回测中使用了当时实际上不可得的数据，导致业绩虚高。在另类数据中，这一问题尤为隐蔽³⁵。

- **时间戳错位**：很多另类数据（如卫星图像、网页抓取数据）的“事件时间”（Event Time）与“数据可用时间”（Arrival Time/Point-in-Time）存在显著滞后。例如，卫星虽然在周一拍摄了照片，但数据处理完成后传输给基金服务器可能已经是周三。如果在回测中假设周一就能利用该信息交易，就是典型的前视偏差。
- **数据修正**：经济数据或公司财报常会被修正。回测必须使用最初发布的“初值”，而非修正后的“终值”。
- **生存偏差 (Survivorship Bias)**：如果不包含已退市或破产的公司，回测结果将严重偏向乐观。另类数据供应商往往只提供现存公司的历史数据，这也是一个常见陷阱。

8.2 基础设施升级

- **存储架构**：从传统的行列式SQL数据库转向支持非结构化数据的数据湖（Data Lakes，如S3，HDFS）和向量数据库（Vector Databases，用于存储文本Embedding）。这对于支持RAG（检索增强生成）和LLM应用至关重要³⁹。
- **云与本地 (Cloud vs. On-prem)**：这是一个经典的权衡。对于探索性研究（Research），本地高性能工作站可能更具成本效益且迭代快；但对于需要处理PB级数据清洗和实时交易（Production）的场景，云端的弹性伸缩能力和存储优势不可替代。然而，云端的高昂成本（特别是出口流量费）也是量化团队必须精算的⁴¹。

9. 法律边界、合规与隐私风险

随着另类数据的威力显现，监管机构（如SEC、欧盟委员会）的关注度也在提升。合规已成为量化基金生存的基石。

9.1 重大非公开信息 (MNPI) 风险

SEC执法的核心在于另类数据是否构成内幕信息。如果数据源（如专家网络、私密社交群组）是非法获取的，或者违反了保密义务，使用该数据即违法⁴²。

- **尽职调查 (Due Diligence)**：基金必须对数据供应商进行严格的审查，确认数据来源的合法性，并确保数据已充分脱敏（De-identified）和聚合（Aggregated），无法追溯到具体个人或单一企业。
- **执法趋势**：2025年的SEC执法报告显示，尽管整体案件数量可能因行政变动有所波动，但针对市场操纵、欺诈和内幕交易的“面包黄油”式案件仍是重点。SEC特别关注那些未能建立完善合规流程来审查另类数据来源的基金⁴⁴。

9.2 网络抓取 (Web Scraping) 与版权

- 合法性: 抓取公开网页通常被认为是合法的 (基于HiQ v. LinkedIn等判例), 但前提是遵守 robots.txt协议, 且不侵入需要登录权限的区域。
- AI训练数据: 欧盟《人工智能法案》(EU AI Act) 对用于训练AI模型的数据提出了透明度要求。这增加了量化机构自建大模型的合规成本, 特别是在处理受版权保护的内容 (如新闻、研报) 时⁴⁸。

9.3 隐私计算与联邦学习

面对日益严苛的隐私法规 (如GDPR, CCPA), 联邦学习 (Federated Learning, FL) 提供了一种解决方案。它允许金融机构在不共享原始数据的前提下, 共同训练反洗钱或信用风控模型。结合差分隐私 (Differential Privacy) 技术的Fed-RD框架, 已被证明能在保护隐私的同时, 维持极高的模型精度, 是未来跨机构数据合作的主流方向⁵⁰。

10. 未来展望: 2026年及以后

站在2026年的门槛上, 量化交易正进入一个新的演进阶段。简单的数据套利机会已被挖掘殆尽, 竞争的焦点转向了更高阶的智能。

10.1 从模型到代理: 多模态AI Agents

传统的量化模型是单一任务的预测器。而最新的研究方向是构建“AI交易代理”(AI Trading Agents)。

- 自主与反思: 这些代理 (如TradingAgents框架) 集成了多模态感知能力 (读新闻、看图表、分析价格), 并拥有“自我反思”(Self-Reflection) 模块。它们能像人类交易员一样, 回顾过去的决策, 总结经验教训, 并动态调整策略。
- 实证表现: FinAgent等模型显示, 这种多模态基础代理在模拟交易中的利润表现比传统基准模型高出36%以上⁵³。

10.2 合成数据 (Synthetic Data) 的战略价值

面对极端行情 (黑天鹅) 数据稀缺的问题, 生成式AI (Generative AI) 大显身手。

- 场景生成: 利用GANs或扩散模型 (Diffusion Models) 生成逼真的、包含极端波动的合成市场数据。这不仅用于压力测试, 还可用于训练更鲁棒的强化学习模型⁵⁵。
- 隐私规避: 合成数据不包含真实个人信息, 因此可以绕过隐私法规进行自由共享和研究, 极大地促进了金融AI的研发速度⁵⁷。

10.3 量化与基本面的终极融合 (Quantamental)

Man Group和Two Sigma等顶级机构在2026年的展望中一致认为, 纯粹的量化与纯粹的基本面界限正在消失。未来的主流是“Quantamental”——利用AI处理海量数据的广度, 结合人类专家的深

度逻辑与直觉。在这个新时代，数据不再是稀缺资源，**对数据的因果推理能力 (Causal Reasoning)**才是真正的Alpha之源⁵⁸。

结语

量化交易与事件驱动策略及另类数据的结合，不仅仅是技术的升级，更是投资哲学的进化。它要求投资者具备跨学科的能力：既要懂金融市场的博弈逻辑，又要精通NLP、计算机视觉及图神经网络等硬核技术。随着市场有效性的不断提升，唯有那些能够持续在数据噪音中构建深度认知、并在合规框架内敏捷迭代的机构，方能在这场算法军备竞赛中立于不败之地。

Works cited

1. Event-Driven Investing | Fund Strategy + Examples - Wall Street Prep, accessed January 27, 2026, <https://www.wallstreetprep.com/knowledge/event-driven-investing/>
2. Beyond Mergers: A Diversified Approach to Event-Driven Investment | AB, accessed January 27, 2026, <https://www.alliancebernstein.com/apac/en/institutions/insights/investment-insights/beyond-mergers-a-diversified-approach-to-event-driven-investment.html>
3. Event-Driven Investment Strategies: How Can Investors Capitalize on Market-Moving Opportunities? - InsideArbitrage, accessed January 27, 2026, <https://www.insidearbitrage.com/2024/11/event-driven-investment-strategies/>
4. Quants: How to Improve Earnings Forecasts with Unique Alt Datasets | ExtractAlpha, accessed January 27, 2026, <https://extractalpha.com/2023/03/08/quants-how-to-improve-earnings-forecasts-with-unique-alt-datasets/>
5. How hedge funds use satellite images to beat Wall Street—and Main Street - Haas News, accessed January 27, 2026, <https://newsroom.haas.berkeley.edu/how-hedge-funds-use-satellite-images-to-beat-wall-street-and-main-street/>
6. How Alternative Data Providers Are Shaping the Next Era of Market Intelligence - Aura, accessed January 27, 2026, <https://blog.getaura.ai/alternative-data-providers>
7. Hedge funds gear up for a 2025 alternative dataset 'budget boom' - Hedgeweek, accessed January 27, 2026, <https://www.hedgeweek.com/hedge-funds-gear-up-for-a-2025-alternative-data-set-budget-boom/>
8. The Alternative Data Market: An Overview (New Figures as of 2024) - TenderAlpha Blog, accessed January 27, 2026, <https://www.tenderalpha.com/blog/post/quantitative-analysis/alternative-data-market-overview-new-figures-as-of-2024>
9. What Is Alternative Data and Why Is It Changing Finance? | Built In, accessed January 27, 2026, <https://builtin.com/articles/alternative-data>

10. LLM vs BERT for Sentiment Classification: Who Wins? | by Asim Adnan Eijaz | Medium, accessed January 27, 2026, <https://medium.com/@asimadnan/llm-vs-bert-for-sentiment-classification-who-wins-1c2995044187>
11. Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models - MDPI, accessed January 27, 2026, <https://www.mdpi.com/2227-7072/13/2/75>
12. Reasoning or Overthinking: Evaluating Large Language Models on Financial Sentiment Analysis - arXiv, accessed January 27, 2026, <https://arxiv.org/html/2506.04574v1>
13. Benchmarking Large Language Models for Target-Based Financial Sentiment Analysis - CLiC-it 2025, accessed January 27, 2026, https://clic2025.unica.it/wp-content/uploads/2025/09/73_main_long.pdf
14. Using NLP to unlock a treasure trove of alternative data | CFA Institute, accessed January 27, 2026, <https://www.cfainstitute.org/insights/articles/using-nlp-to-unlock-treasure-trove-of-alternative-data>
15. Predicting M&A targets using news sentiment and topic detection - IDEAS/RePEc, accessed January 27, 2026, <https://ideas.repec.org/a/eee/tefoso/v201y2024ics0040162524000660.html>
16. Predicting M&A targets using news sentiment and topic detection | RUN, accessed January 27, 2026, https://run.unl.pt/bitstream/10362/163548/1/Predicting_M_A_targets_using_news_sentiment_and_topic_detection.pdf
17. Alternative Data in Quantitative Strategies: Use Cases - Informa Connect, accessed January 27, 2026, <https://informaconnect.com/alternative-data-in-quantitative-strategies-use-cases/>
18. I've heard hedge funds use satellite images to monitor fluctuations in parking lot traffic for chain stores. Is this true? : r/gis - Reddit, accessed January 27, 2026, https://www.reddit.com/r/gis/comments/1azazam/ive_heard_hedge_funds_use_satellite_images_to/
19. Eye in the sky: How satellite imagery can be used to solve real business problems - Medium, accessed January 27, 2026, <https://medium.com/ideachain/eye-in-the-sky-how-satellite-imagery-can-be-used-to-solve-real-business-problems-13b5d5bb6098>
20. This Startup Makes Money from Oil Tank Shadows - Digital Innovation and Transformation, accessed January 27, 2026, <https://d3.harvard.edu/platform-digit/submission/this-startup-makes-money-from-oil-tank-shadows/>
21. Stalking Farms: Predicting Agricultural Yield with Satellite Imagery - CS231n, accessed January 27, 2026, <https://cs231n.stanford.edu/2024/papers/stalking-farms-predicting-agricultural-yield-with-satellite-imag.pdf>
22. Agriculture Case Studies Using Satellite Data | Pixalytics LtdPixalytics Ltd,

- accessed January 27, 2026, <https://www.pixalytics.com/agriculture-case-satellite/>
23. Quantitative Trading: Data-Driven Strategies - LuxAlgo, accessed January 27, 2026, <https://www.luxalgo.com/blog/quantitative-trading-data-driven-strategies/>
 24. 5 data cleaning techniques for high-impact financial analysis - Finance Alliance, accessed January 27, 2026, <https://www.financealliance.io/data-cleaning-techniques/>
 25. Financial Services Data Analytics & Transaction Data Cleansing - Alkami Technology, accessed January 27, 2026, <https://www.alkami.com/blog/an-overview-of-financial-services-data-analytics-transaction-data-cleansing-for-financial-institutions/>
 26. Unlocking Alpha With Alternative Data - CFM - Capital Fund Management, accessed January 27, 2026, <https://www.cfm.com/unlocking-alpha-with-alternative-data/>
 27. Supply Chain Network Analytics with Knowledge Graphs - ArcGIS StoryMaps, accessed January 27, 2026, <https://storymaps.arcgis.com/stories/8dcb1b0587124657bebe20b57ba7310f>
 28. FinReflectKG: Agentic Construction and Evaluation of Financial Knowledge Graphs - arXiv, accessed January 27, 2026, <https://arxiv.org/html/2508.17906v2>
 29. Intricate Supply Chain Demand Forecasting Based on Graph Convolution Network - MDPI, accessed January 27, 2026, <https://www.mdpi.com/2071-1050/16/21/9608>
 30. Design and Performance Prediction for Supply Chain Systems with Graphical Structures, accessed January 27, 2026, <https://dukespace.lib.duke.edu/items/2816bc60-c79f-467b-abd4-80498e98b0ae>
 31. Time Series Isn't Enough: How Graph Neural Networks Change Demand Forecasting, accessed January 27, 2026, <https://towardsdatascience.com/time-series-isnt-enough-how-graph-neural-networks-change-demand-forecasting/>
 32. Supply Chain Link Prediction on Uncertain Knowledge Graph - University of Cambridge, accessed January 27, 2026, <https://www.repository.cam.ac.uk/bitstreams/826e207c-4a31-446b-800c-9ee70150b4e5/download>
 33. Supply Chain Link Prediction on Uncertain Knowledge Graph, accessed January 27, 2026, https://kdd.org/exploration_files/p124-AI4Manufacturing_paper5.pdf
 34. Supply Chain Financial Fraud Detection Based on Graph Neural Network and Knowledge Graph, accessed January 27, 2026, <https://hrcak.srce.hr/file/465286>
 35. Alternative data at investment management firms: From discovery to integration - Deloitte, accessed January 27, 2026, <https://www.deloitte.com/us/en/insights/industry/financial-services/alternative-data-for-investors-from-discovery-to-institutionalization.html>
 36. Look-Ahead Bias In Backtests And How To Detect It | by Michael Harris | Medium, accessed January 27, 2026, <https://mikeharrisny.medium.com/look-ahead-bias-in-backtests-and-how-to-detect-it-ad5e42d97879>
 37. Backtesting Traps: Common Errors to Avoid - LuxAlgo, accessed January 27,

- 2026, <https://www.luxalgo.com/blog/backtesting-traps-common-errors-to-avoid/>
38. Problems in Backtesting - CFA, FRM, and Actuarial Exams Study Notes - AnalystPrep, accessed January 27, 2026, <https://analystprep.com/study-notes/cfa-level-2/problems-in-backtesting/>
39. AI-ready Infrastructure for Quantitative Trading | Pure Storage, accessed January 27, 2026, <https://www.purestorage.com/content/dam/pdf/en/solution-briefs/sb-airi-quant-trading.pdf>
40. How to set up a Quantitative Trading Market Data Repository | by Mario Emmanuel - Medium, accessed January 27, 2026, <https://medium.com/data-science/a-quantitative-trading-market-data-infrastructure-4ca7578e0291>
41. Cloud or Local: Where to Run Your Quant Trading? - QuantRocket, accessed January 27, 2026, <https://www.quantrocket.com/blog/cloud-vs-local/>
42. Data Compliance and MNPI: Protect Information with PBAC - PlainID, accessed January 27, 2026, <https://www.plainid.com/data-compliance-and-mnpi/>
43. Key Considerations for Alternative Data and AI Vendors to Investment Firms: Demonstrating Compliance in the Face of an Evolving Regulatory Environment | Lowenstein Sandler LLP, accessed January 27, 2026, <https://www.lowenstein.com/news-insights/publications/articles/key-considerations-for-alternative-data-and-ai-vendors-to-investment-firms-demonstrating-compliance-in-the-face-of-an-evolving-regulatory-environment>
44. SEC Focused On Fraud As Actions Markedly Declined In 2025 - Covington & Burling LLP, accessed January 27, 2026, <https://www.cov.com/-/media/files/corporate/publications/2025/11/sec-focused-on-fraud-as-actions-markedly-declined-in-2025.pdf>
45. Securities Enforcement 2025 Mid-Year Update - Gibson Dunn, accessed January 27, 2026, <https://www.gibsondunn.com/securities-enforcement-2025-mid-year-update/>
46. WMHW Alert: SEC Enforcement Trends in the First Six Months, accessed January 27, 2026, <https://wmhwlaw.com/2025/10/17/wmhw-alert-sec-enforcement-trends-in-the-first-six-months/>
47. April 26, 2022 Investment Adviser MNPI Compliance Issues I. Introduction The Division of Examinations ("EXAMS")* is issuing - SEC.gov, accessed January 27, 2026, <https://www.sec.gov/files/code-ethics-risk-alert.pdf>
48. The EU's AI Act: Review and What It Means for EU and Non-EU Companies, accessed January 27, 2026, <https://www.pillsburylaw.com/en/news-and-insights/eu-ai-act.html>
49. The European Union is still caught in an AI copyright bind - Bruegel, accessed January 27, 2026, <https://www.bruegel.org/analysis/european-union-still-caught-ai-copyright-bind>
50. Privacy-Preserving Machine Learning in Financial Customer Data: Trade-Offs Between Accuracy, Security, And Personalization - AJHSSR, accessed January 27, 2026,

- <https://www.ajhssr.com/privacy-preserving-machine-learning-in-financial-customer-data-trade-offs-between-accuracy-security-and-personalization/>
51. Introduction to Privacy-Preserving Techniques in Financial AI - Wissen, accessed January 27, 2026, <https://www.wissen.com/blog/introduction-to-privacy-preserving-techniques-in-financial-ai>
 52. Fed-RD: Privacy-Preserving Federated Learning for Financial Crime Detection - arXiv, accessed January 27, 2026, <https://arxiv.org/html/2408.01609v1>
 53. TradingAgents: Multi-Agents LLM Financial Trading Framework - arXiv, accessed January 27, 2026, <https://arxiv.org/pdf/2412.20138>
 54. [PDF] A Multimodal Foundation Agent for Financial Trading: Tool-Augmented, Diversified, and Generalist | Semantic Scholar, accessed January 27, 2026, <https://www.semanticscholar.org/paper/A-Multimodal-Foundation-Agent-for-Financial-and-Zhang-Zhao/c86a70ff639707e647da3a429fe8e1e5c04415f5>
 55. Full article: Generation of synthetic financial time series by diffusion models, accessed January 27, 2026, <https://www.tandfonline.com/doi/full/10.1080/14697688.2025.2528697>
 56. stefan-jansen/synthetic-data-for-finance: Material for QuantUniversity talk on Synthetic Data Generation for Finance. - GitHub, accessed January 27, 2026, <https://github.com/stefan-jansen/synthetic-data-for-finance>
 57. Generating synthetic data in finance: opportunities, challenges and pitfalls - J.P. Morgan, accessed January 27, 2026, <https://www.jpmorgan.com/content/dam/jpm/cib/complex/content/technology/ai-research-publications/pdf-8.pdf>
 58. AI in Investment Management: 2026 Outlook (Part II) - Two Sigma, accessed January 27, 2026, <https://www.twosigma.com/articles/ai-in-investment-management-2026-outlook-part-ii/>
 59. Goldman Sachs Asset Management Releases 2026 Investment Outlook, accessed January 27, 2026, <https://am.gs.com/en-us/advisors/news/press-release/2025/goldman-sachs-asset-management-releases-2026-investment-outlook>
 60. Stock Market Outlook 2026: Political Risks Loom | Morgan Stanley, accessed January 27, 2026, <https://www.morganstanley.com/insights/articles/2026-market-optimism-and-risks>