

Abstract

This project develops a four-agent, Fundamental-led equity research framework that combines Fundamental, Valuation, Sentiment, and Macro evidence. Across two 52-day windows, the strategy reduced downside risk in the conflict period but underperformed Buy and Hold in both windows.

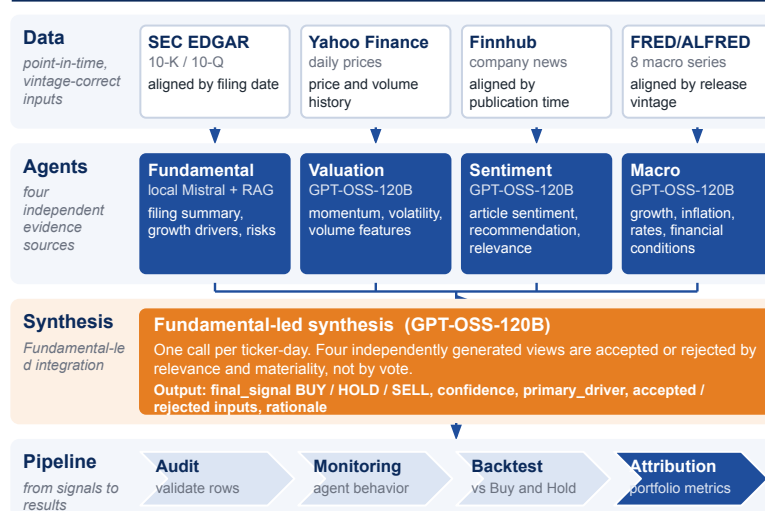
Architecture Evolution

- Free debate (inherited)**
Any disagreement opened an AutoGen group chat among the three agents.
- Narrowed trigger**
21 of 208 CAT and CVX decisions changed, all toward BUY, without new evidence.
- Hierarchical debate**
Fundamental leads, peers challenge the base case; 82% of days retain it.
- Fundamental-led synthesis**
Four independent sources; one Fundamental-led call accepts or rejects each.

Experimental Design

- Universe**
20 large-cap U.S. equities across sectors
- Backtest Periods**
Window 1, conflict: 2026-02-11 to 2026-04-27
Window 2, calm: 2025-05-28 to 2025-08-11
52 trading days each
- Data and Point-in-Time**
Fundamental SEC filings by filing date
News Finnhub articles by publication time
Macro FRED / ALFRED series by release vintage
- Trading Signal**
Daily BUY / HOLD / SELL per ticker
Formed on day t and executed on day $t + 1$
- Portfolio Construction**
Fixed initial-weight sleeves, no cross-stock reallocation
5 bps transaction cost per position change
Benchmark: the same 20 stocks held continuously

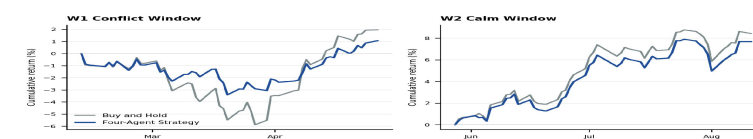
System Architecture



Macro Agent

The Macro Agent was integrated as a fourth evidence source. Indicators are categorized into four pillars: Growth (GDP, Unemployment), Inflation (CPI), Rates (Fed Funds, Treasury Yields, Yield Curve), and Financial Conditions (VIX, Credit Spreads). Slow-moving economic indicators are sourced from ALFRED vintages to ensure point-in-time accuracy, while daily market and policy indicators are pulled directly from FRED. Fixed, interpretable thresholds are used for structural baseline conditions; for dynamic signals, the agent compares current observations against expanding historical 25th and 75th percentiles. The four component scores are weighted for sector-specific sensitivity. Across 18 prompt variants, scores-only configurations significantly outperformed features-only models, which hallucinated irrelevant company-specific information.

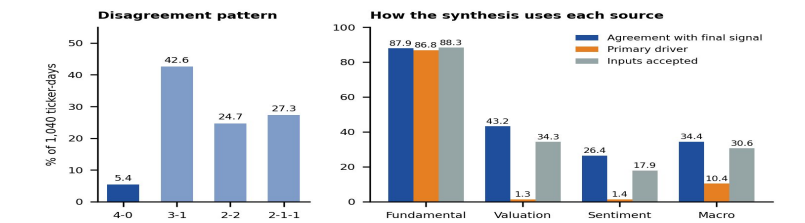
Cumulative Performance vs Buy and Hold



Key Results

Metric	W1 B&H	W1 Strategy	W2 B&H	W2 Strategy
Total return	+1.96%	+1.06%	+8.42%	+7.68%
Annualized volatility	11.08%	7.34%	8.77%	8.33%
Sharpe (rf = 5%)	0.452	0.051	3.947	3.748
Max drawdown	-5.88%	-3.41%	-2.68%	-2.73%
Avg gross exposure	100%	92.85%	100%	92.08%
Trades	0	175	0	113

Synthesis Behavior



The sources rarely agree: full four-way agreement occurred on only 5.4% of days. The final signal tracks the Fundamental Agent closely, departing from its independent view on 126 days, 12.1% of ticker-days. Of the 137 days with a non-Fundamental primary driver, 108 (78.8%) were the Macro Agent; the Sentiment Agent's inputs were rejected on 82.1% of days. Final signals were BUY on 85.3% of days: the bullish skew is reduced, not removed.

Limitations

Portfolio returns were below Buy and Hold in both windows, 1.06% versus 1.96% in W1 and 7.68% versus 8.42% in W2. The reduction in volatility and drawdown in the conflict window did not translate into better risk-adjusted performance because exposure changes were not consistently well timed across stocks. The evaluation covers only two 52-trading-day windows, the 126 synthesis overrides have not yet been linked to realized forward returns, and the Macro Agent's smoothing can delay its response to sudden shocks.