

LLM Sentiment Equity Analysis

Financial News & Filings to Investment Signals

Ethan Davis, Ross Cunningham, Akash Hari

Our Project

- Financial markets react to **new information**, much of which is **text** (news, earnings, SEC filings).
- Show how **LLMs can convert unstructured text into tradable signals**, and evaluate which designs remain robust once we add realistic timing and trading frictions.

01. Use the **FNSPID financial news dataset** and **FinBERT** to turn headlines into daily sentiment signals and test whether they improve short-horizon equity strategies.

02. Build a **SEC filings pipeline** (10-K/10-Q/8-K) where an LLM extracts structured features (sentiment, risk, tone) that feed into a machine-learning model for return prediction.

Wikipedia / Books / Webtext



BERT Model



Fine Tune on
Financial Text

FINBERT Model

Financial
News Article
or Company
Record



FINBERT
Model



Pos/Neu/Neg
Probabilities

What is FINBERT?

BERT (Bidirectional Encoder Representations from Transformers)

- Transformer model that reads text **both left-to-right and right-to-left**
- Pretrained on large generic text corpora (e.g. Wikipedia, BooksCorpus) and then **fine-tuned** for specific tasks

FinBERT is A BERT model fine-tuned on financial documents (reports, earnings transcripts, news). Specialised vocabulary and weights so it understands finance-specific language

- Input: raw text from a news article linked to a stock
- Output: probabilities that the text is positive, neutral, or negative from an investor's point of view

Datasets

- Financial News and Stock Price Integration Dataset combining stock prices and financial news for S&P 500 companies
- Scale: 29.7 million stock price records and 15.7 million news articles for 4,775 firms from 1999 to 2023, collected from four major market news sites
 - In this project we work with a filtered subset: 40 large-cap S&P-like tickers from 2010–2024, resulting in ~19k news articles after slicing and cleaning.
- Structure: Each news item is time-aligned with ticker and date; we construct a unified news.csv with **post_id**, **ticker**, **created_at**, **date**, text before running FinBERT
- Our use: we use FNSPID as a source of historical prices and raw news text, on top of which we compute FinBERT sentiment scores and build a daily price–sentiment panel for forecasting.

Data Preprocessing Steps

01.

News Construction

Filtered raw FNSPID news to a focused large cap universe and built a clean article level dataset.



- Slice 22GB NASDAQ file by ticker and date with chunked loading
- Keep post_id, ticker, timestamp, date and combined title plus body
- Clean text (lowercase, remove URLs and extra whitespace)
- Save unified news.csv for NLP

02.

NLP + Daily Sentiment

Apply FinBERT to each article and roll results up to daily sentiment features.



- Get FinBERT p_pos, p_neg, p_neu for every article
- Derive net sentiment (pos - neg) and intensity (|pos - neg|)
- Count strongly positive and strongly negative stories
- Aggregate by ticker-day into net, intensity, disagreement, volume and 5 day surprise

03.

Price + Returns

Builds in each stock's open, close, adj close, and volume.



- Load daily OHLCV for the same ticker universe
- Use adjusted close to compute log returns and forward one day returns
- Create direction labels and five day realised volatility
- Merge price and sentiment on ticker and date into a single modeling panel

Feature Engineering

Sentiment:

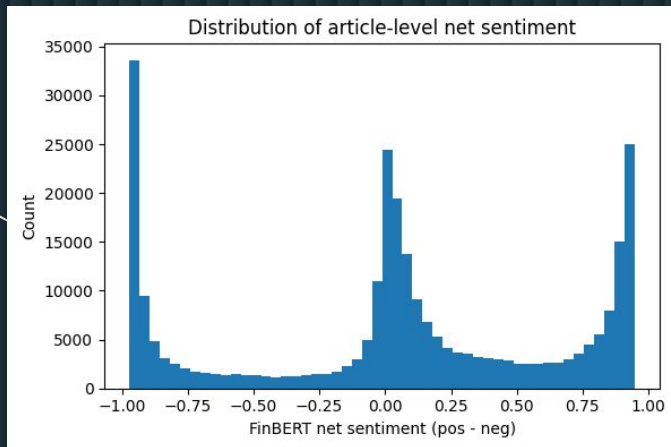
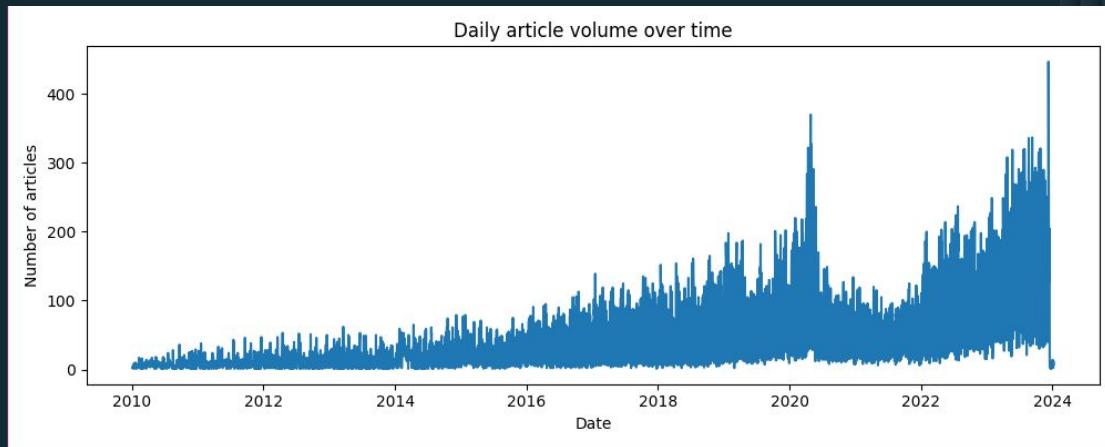
sent_net:	mean FinBERT net sentiment (pos-neg) per ticker-day.
sent_intensity:	mean pos-neg (strength of tone).
sent_disagreement:	variance of article-level net sentiment.
sent_volume:	number of articles.
sent_surprise_5d:	deviation of today's sentiment from 5-day average.
strong_pos_ratio / strong_neg_ratio:	fraction of strongly positive / negative articles.

Market Variables:

ret	daily log return
ret_fwd1	next-day log return
direction_fwd1:	dummy variable: 1 if <code>ret_fwd1</code> > 0
moving_avgs	50 and 150 day moving averages
vol_real_5d	5-day rolling realised volatility

Exploratory Analysis: Sentiment

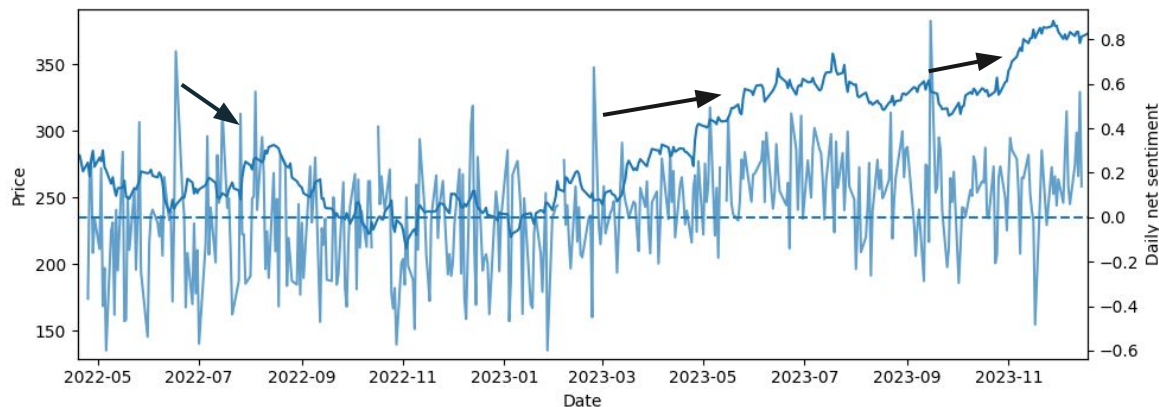
- Large Spikes in Dataset Article Volume often centered around key economic events (COVID-19 Spike)
- Net Sentiment Scores are more focused on edges (-1,1, Pure Sentiment) rather than unexplained opinions



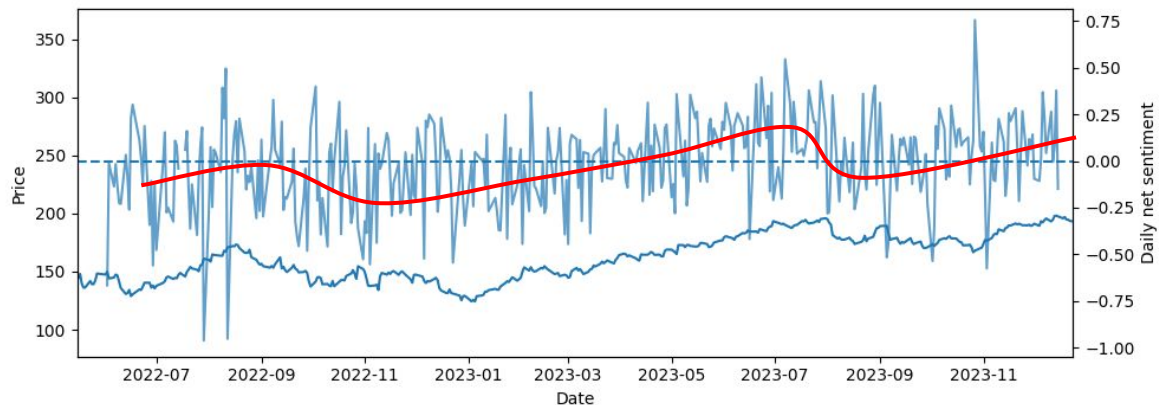
Exploratory Analysis: Sentiment vs Stock Price

- Subtle Correlations between price and news sentiment
- Sentiment Spikes triggering large price swings within a couple months
- Could this be earning's release related? Or just a coincidence

MSFT: Price and daily net sentiment over time



AAPL: Price and daily net sentiment over time

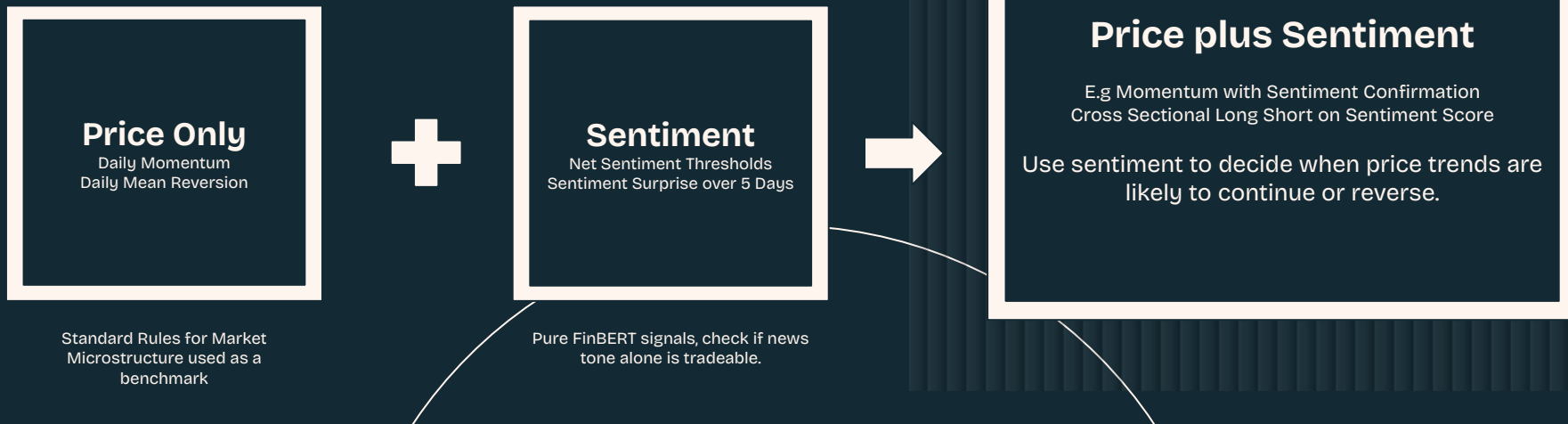


Strategy Breakdown

We want to measure the gap between news sentiment and realised returns and see which trading strategies actually benefit from sentiment scores.

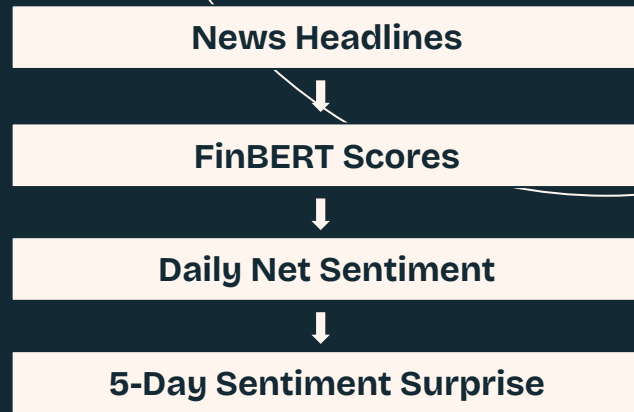
So where do we start?

We first build simple baseline strategies, then add sentiment scores on top to see how much extra information FinBERT provides.



Raw Sentiment -> Surprise Feature Example

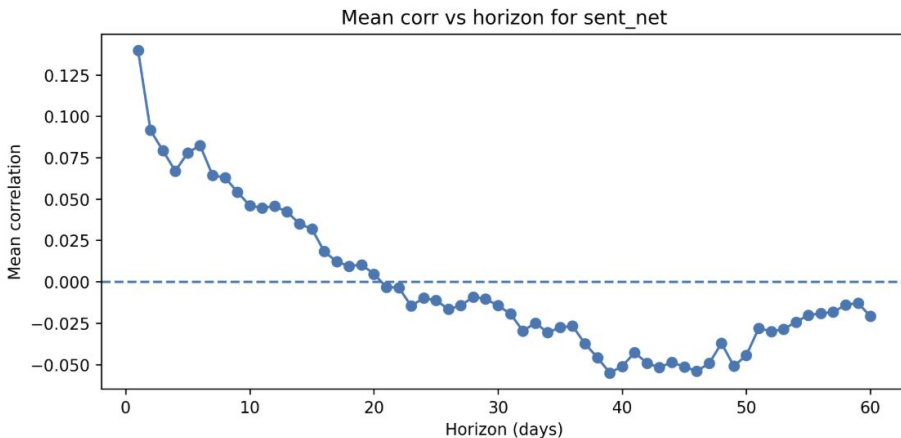
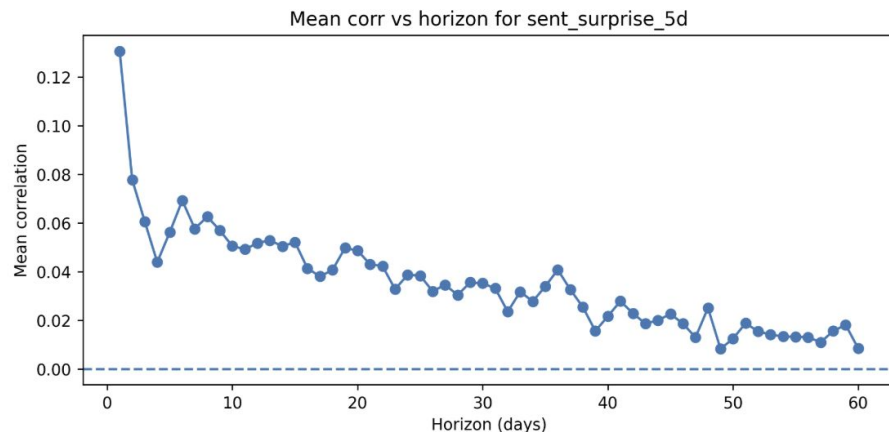
- For each stock we pull all firm specific headlines and run them through FinBERT.
- Each article gives positive and negative probabilities; we combine these into a **net sentiment** for that article.
- We aggregate articles by stock and date to get a **daily net sentiment score**.
- Finally we build **sentiment surprise**:
 - average net sentiment over the last five days
 - compare it to the stock's own recent history using a rolling mean and volatility
- This gives one clean number per stock per day that tells us when the recent news tone is unusually good or unusually bad.



Date	# Articles	Avg Pos Prob	Avg Neg Prob	Net Sentiment	5-day Sentiment Surprise
2025-01-02	4	0.72	0.14	+0.58	+1.3 σ
2025-01-03	3	0.65	0.18	+0.47	+1.1 σ
2025-01-06	5	0.60	0.22	+0.38	+0.9 σ
2025-01-07	2	0.55	0.25	+0.30	+0.6 σ
2025-01-08	3	0.50	0.30	+0.20	+0.2 σ

Picking The Horizon:

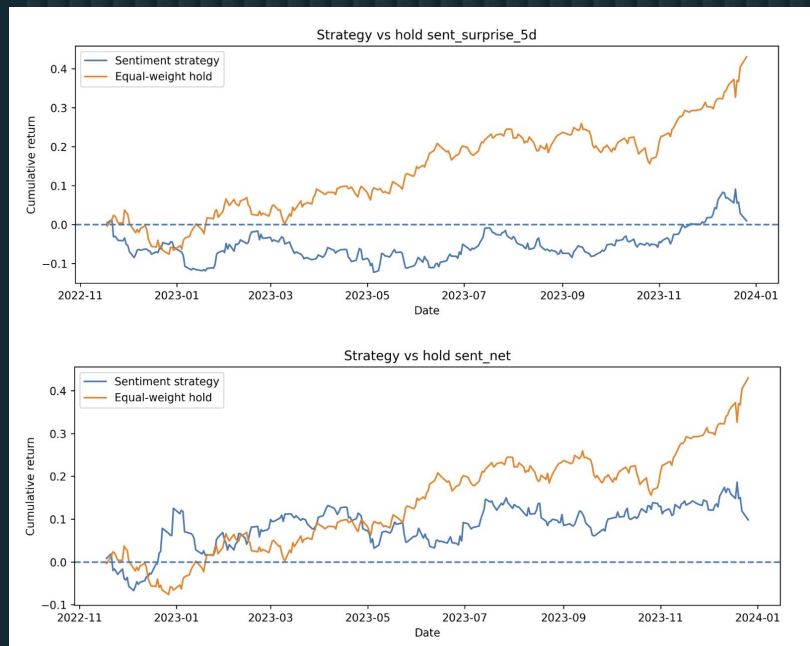
- We tested how well sentiment surprise predicts returns over horizons from **one to sixty trading days**.
- Correlations are highest at very short horizons and **decay quickly** as we move further out.
- Across stocks the signal is strongest between **one and five days**, after that the relationship becomes weak and noisy.
- Based on this, we design trading rules that react to **recent sentiment** and hold positions for about **five days** rather than for months.



Price Plus Sentiment: Backtest 2018 to 2023

- Strategy: daily cross sectional long short
 - Long stocks with very positive sentiment feature
 - Short stocks with very negative sentiment feature
 - Code Built to change sentiment feature easily to allow for lots of different tests

- On 2018 to 2023 data the **price plus sentiment** version improves both accuracy and Sharpe ratio compared to price only rules.
- This suggests that FinBERT sentiment is adding **genuine extra information** on top of simple price based signals.
- At this stage we have not yet included realistic trading costs.



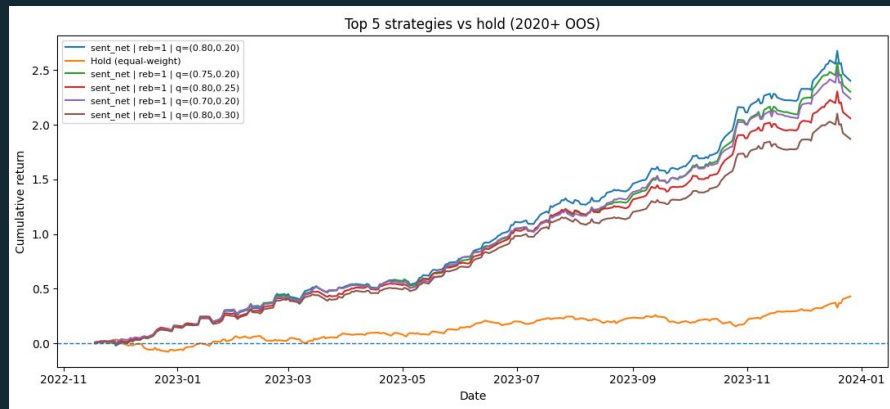
New 2024–2025 Data: Can We Still Trade It?



- To avoid reusing the same sample we pulled new headlines for 2024 and 2025 for our stock set using Finnhub and the FNSPID tools.
- We scored these articles with FinBERT using the same model and processing as in the original sample.
- We then applied the same long short strategy but now in a realistic implementation:
 - build sentiment on day t , trade at the close of t
 - earn forward close to close returns
 - include trading costs based on daily turnover.
- On this fresh period the long short strategy loses money and underperforms both the equal weight universe and the general market.
- The signal is not strong enough to survive high turnover and realistic costs in a pure long short set up.

Learnings and What Actually Works

- Our original long short results looked strong in FINSPID, but the 2025 out of sample test shows the edge is much less stable once we change the news source and shorten the window.
- The most important data reality is coverage: FINSPID has dense daily article flow, while our 2025 Finnhub build captures far fewer articles, with many ticker days having no news, so the sentiment signal is often missing or noisy.
- In practice this shifts the usable sample dramatically, from roughly 94% usable ticker days in FINSPID to about 13% in 2025, which reduces signal reliability and makes strategy performance much more sensitive to a few weeks of market moves.
- The takeaway is that FinBERT is more reliable as a filter or tilt than as a full long short book in the 2025 panel, because short legs and frequent rebalancing amplify noise when coverage is thin.
- Most realistic implementation: use sentiment to either drop the worst names or lightly overweight the strongest names, and rebalance less frequently to reduce turnover and transaction costs.



Predicting Stock Returns Using LLM-Derived Insights from SEC Filings



- Investors rely heavily on qualitative information in SEC filings (10-K, 10-Q, 8-K)
- Traditional quant signals under-use textual sentiment
- Recent advances in LLMs allow structured extraction of financial signals from long filings
- Goal: Build a pipeline that converts SEC filings into LLM features into ML signal to generate a trading strategy

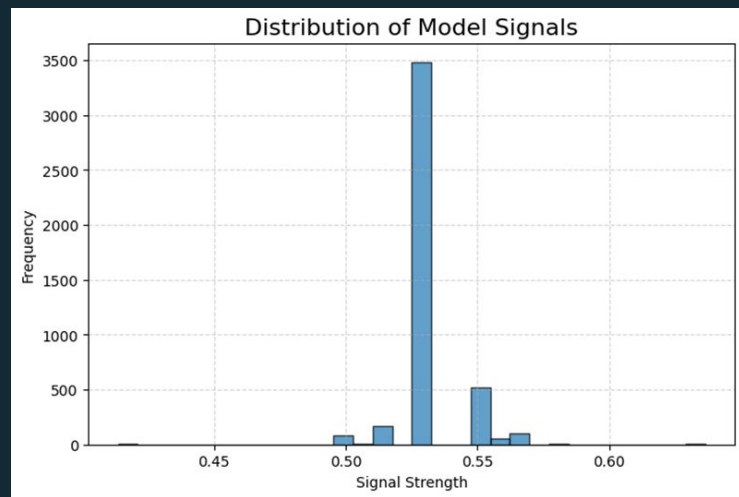
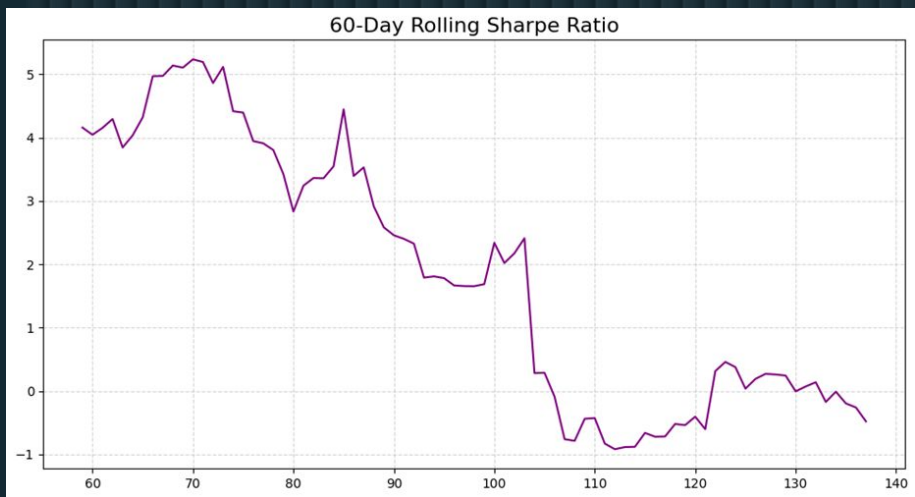
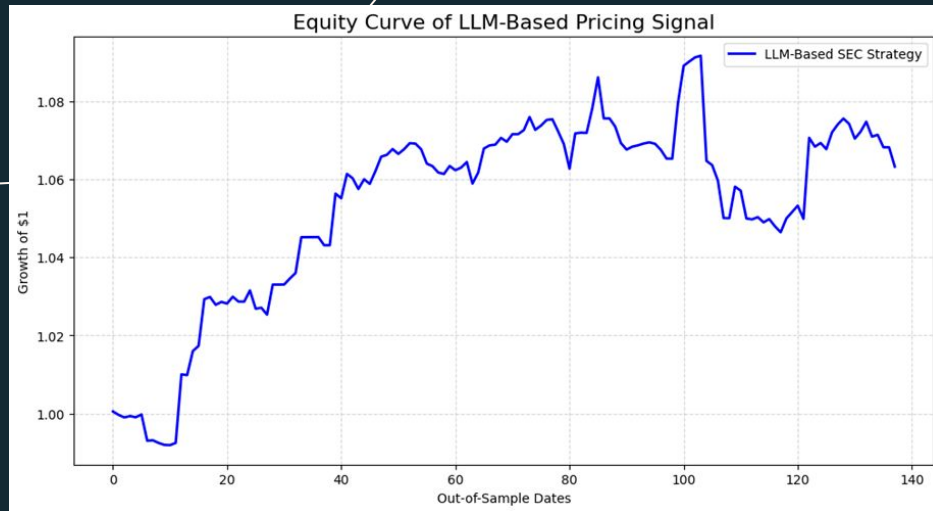
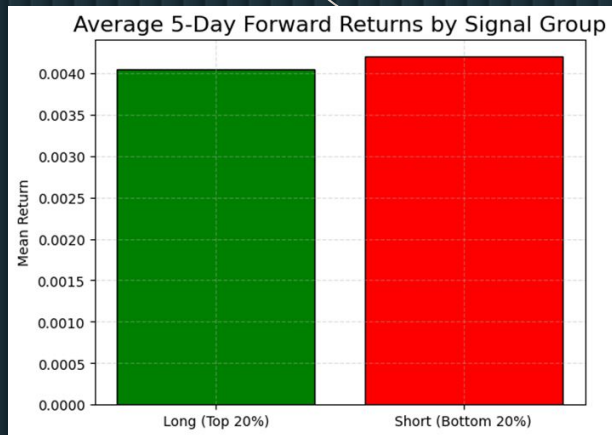
Data Pipeline Overview

- 1. Ticker Universe: S&P500 pulled from Slickcharts
- 2. SEC Filing Scraper: Downloads historical 10-K, 10-Q, and 8-K filings
- 3. Text processing: Extracts text sections from SEC HTML filings
- 4. LLM Analysis: GPT analyzes text and returns structured JSON features
- 5. Price Labeling: Aligns filing dates to market data and computes future returns (r_1 , r_5 , r_{21})
- 6. Machine Learning Model: Predicts direction of returns
- 7. Backtesting Engine: Evaluates signals out-of-sample

Machine Learning Model - Predict Stock Return Direction

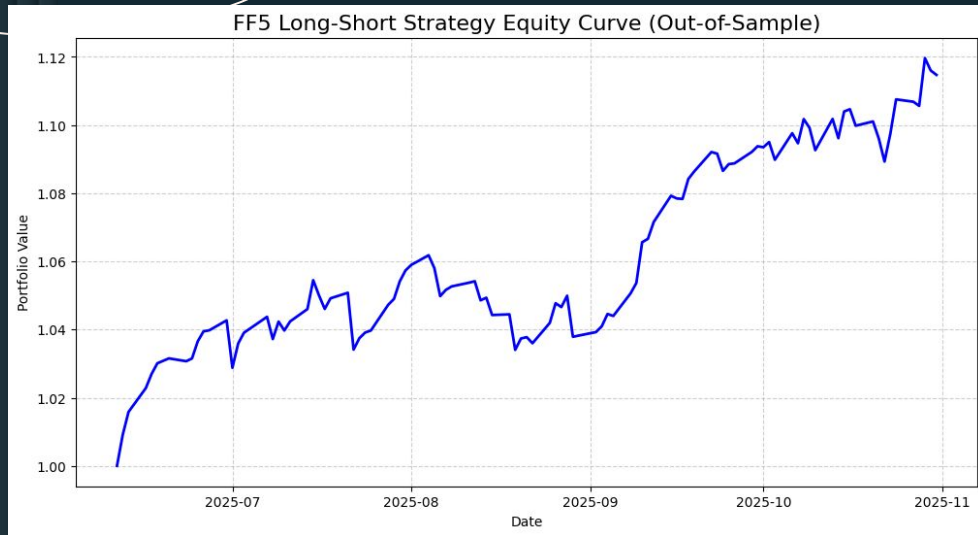
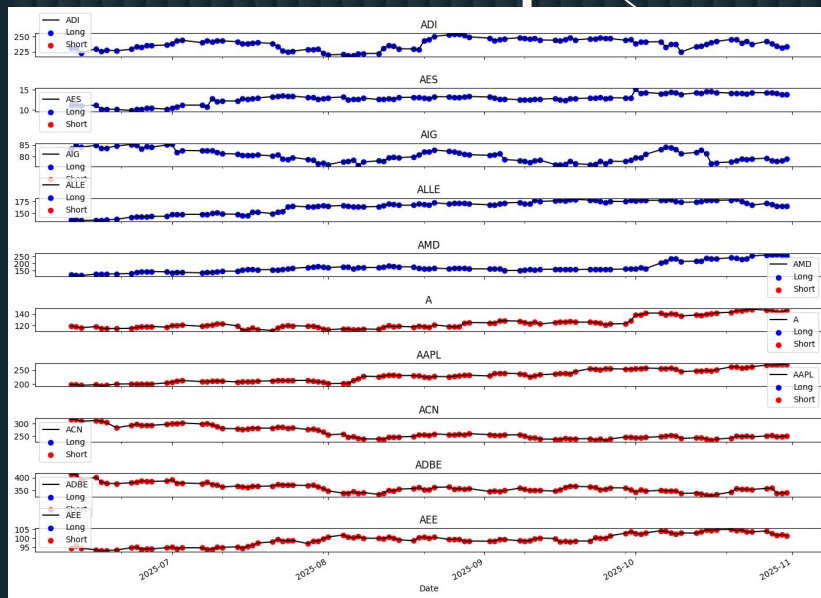
- Model: LightGBM - handles nonlinear, noisy financial data
 - fast scalable, and resistant to overfitting
 - learns interactions between sentiment, tone and risk indicators
- Input: LLM features (sentiment, risk, tone)
- Target: $r_5 > 0$ (binary classification)
- Train/test split based on time (70/30)
- Probabilities naturally map to long/short portfolio weights
- Future upgrades: XGBoost, transformer embeddings

Results



Factor Comparison

- Test results against traditional finance pricing signal strategy (e.g. Fama & French 5 factor model)



Strategy	FF5 Signal	LLM Signal
Cumulative Return	11.47%	6.32%
Annualized Return	31.48%	12.44%
Annualized Vol	7.86%	7.40%
Sharpe Ratio	4.00	1.55
Max Drawdown	2.78%	4.51%

Discussion & Limitations

Positive results showcase motivation to combine three models

- **LLM Filing Model** — slow-moving, fundamentals-based linguistic signals
- **LLM News Model** — high-frequency sentiment shocks and narrative drift
- **FF5 Factors** — structural risk premia explaining cross-sectional returns

Current Limitations

- **Sparse event frequency:** Signals only occur on filing dates, limiting trading opportunities.
- **LLM constraints:** Token limits and cost restrict full-document processing.
- **Partial text coverage:** Only the first ~9,000 characters of each filing are analyzed.
- **Simple linguistic features:** Primarily sentiment and tone—missing deeper semantic structure.

Future Improvements

- **Full-document embeddings** using transformer-based encoders.
- **Chunking + hierarchical summarization** to cover 100% of each filing.
- **More advanced ML models:** XGBoost, ensembles, hybrid LLM–tree systems.
- **Market frictions:** Incorporate transaction costs and realistic execution.
- **Alternative prediction horizons** (1-day, 20-day, long-window sentiment drift).