# News Sentiment Analysis and Summarization for Financial Markets

*ORIE 5253 Asset Management Seminar Report*

Kexin Deng*[1], Weihang Lin†[2], and Sirui Zhao‡[3]

[1,2,3]Cornell University, ORIE, Financial Engineering§

November 2025

**Abstract**

We build an end-to-end system that ingests company-specific financial news, produces article-level sentiment scores using a domain-specific language model (FinBERT), aggregates these into daily stock-level sentiment time series, and presents the output in an investor-friendly dashboard. The platform combines sentiment and price co-movements, simple Buy/Hold/Sell regimes, and a large-language-model (LLM) based news summary agent. Focusing on S&P 500 constituents, we describe the data construction, modeling pipeline, and interface design, and we report descriptive empirical results on the behavior of the sentiment signal and its relation to returns. We also outline a roadmap toward portfolio-level applications and deeper integration with fundamental data.

## 1 Introduction

In modern equity markets, prices adjust continuously to new information, but the reasons behind these moves are dispersed across a vast and noisy stream of financial news. Institutional investors often rely on expensive terminals to digest this flow, while retail investors must work with fragmented news feeds from brokers and finance portals. In both settings, there is a gap between raw headlines and actionable, structured insight.

This project aims to narrow that gap by building an AI-driven platform for news sentiment analysis and summarization in financial markets. The system ingests company-specific news in near real time, classifies the sentiment of each article using FinBERT, aggregates those signals into daily sentiment series at the stock level, and presents them via a set of panels that combine:

- sentiment and price co-movements,
- simple Buy/Hold/Sell regimes based on smoothed sentiment, and
- natural-language news summaries generated by an LLM-based agent.

We focus on large-cap U.S. equities, using the S&P 500 as the initial universe. Our design is guided by three principles:

*Email:kd537@cornell.edu

†Email:wl769@cornell.edu

‡Email:sz695@cornell.edu

§We are grateful to Professor Andrew Chin and the other committee members for their valuable suggestions.

1. **Domain-specific NLP**: rely on models trained on financial text rather than generic corpora.
2. **Transparent aggregation**: apply clear, interpretable formulas to move from article-level scores to daily signals.
3. **Investor-friendly presentation**: deliver outputs in formats a portfolio manager can absorb in seconds.

These considerations lead to three core research questions:

(a) *Aggregation:* How can we aggregate heterogeneous news items into a single, meaningful sentiment time series for each stock?

(b) *Performance:* How stable are these sentiment signals, and how do they relate qualitatively to price dynamics?

(c) *Recommendations:* How should we translate sentiment into intuitive, decision-oriented views such as Buy/Hold/Sell bands and concise narratives?

The rest of the report is organized as follows. Section 2 reviews related literature on investor sentiment, financial NLP, and LLM-based summarization. Section 3 describes the data and sample construction. Section 4 details the modeling pipeline. Section 5 presents empirical results and qualitative insights. Section 7 discusses limitations and future research directions, and Section 9 concludes.

# 2 Literature Review

## 2.1 Investor Sentiment and News

A large literature links investor sentiment to asset prices. Baker and Wurgler (2006) show that broad sentiment affects the cross-section of returns, particularly for young, volatile, and hard-to-arbitrage stocks. At the market level, Tetlock (2007) construct a daily pessimism index from *Wall Street Journal* columns and find that high pessimism predicts temporarily depressed market returns, with prices tending to mean-revert as sentiment normalizes. Stambaugh et al. (2012) further document that many well-known anomalies are stronger following high-sentiment periods, consistent with mispricing that is later corrected when limits to arbitrage bind.

Subsequent work moves from index-level to firm-level news. Using textual analysis of newswires, Boudoukh et al. (2013) show that only a small subset of firm-specific news items is associated with large price moves, while much of the flow is quickly absorbed with little impact. Media coverage itself also shapes attention and pricing: Fang and Peress (2009) find that stocks with low media coverage earn higher average returns, and Da et al. (2011) use Google search volume as a direct measure of investor attention, showing that high attention predicts temporarily elevated prices and subsequent reversals.

More recent work expands sentiment sources to social media and other alternative platforms. Bollen et al. (2011) construct mood indices from Twitter messages and show that aggregate mood helps forecast broad market returns over short horizons. At the firm level, Ranco et al. (2015) adapt an event-study design to Twitter data and show that sentiment around peaks in tweet volume about Dow Jones stocks predicts short-horizon cumulative abnormal returns. Together, these studies suggest that text-based sentiment captures both broad market tone and fine-grained, event-driven signals.

Taken together, this literature supports three key points that motivate our project:

- News and textual sentiment contain information relevant for short- to medium-term returns.

- Predictive power is often strongest at the stock-specific and event-driven level.
- The impact of sentiment is asymmetric and state-dependent, with negative news often having more persistent effects than positive news.

## 2.2 Financial Sentiment Models and FinBERT

Early sentiment approaches in finance relied on dictionary-based methods or generic bag-of-words classifiers. A key insight of Loughran and McDonald (2011) is that general-purpose sentiment lexicons misclassify many common financial terms (such as "liability" or "capital") as negative, leading to biased scores in 10-K filings. Subsequent work develops domain-specific resources, most notably the Financial PhraseBank of Malo et al. (2014), which provides sentence-level sentiment labels for financial and economic news and has become a standard benchmark for financial sentiment models.

Transformer-based language models changed this landscape. Devlin et al. (2019) introduce BERT, which learns contextual word representations by pre-training on large generic corpora and then fine-tuning on downstream tasks. Building on this idea, Araci (2019) and Yang et al. (2020) independently develop FinBERT variants that are further pre-trained on financial text—including 10-K/10-Q filings, earnings-call transcripts, and analyst reports—and then fine-tuned on datasets like Financial PhraseBank for sentence-level sentiment classification. Across multiple benchmarks, these domain-adapted models outperform dictionary-based methods and generic BERT on financial sentiment tasks.

In parallel, work on large language models for return prediction illustrates how richer language understanding translates into economically meaningful signals. Lopez-Lira and Tang (2023) show that ChatGPT's classifications of news headlines into good, bad, or neutral for a stock predict next-day returns and subsume the information in a commercial news-sentiment score. Xie et al. (2023) analyze zero-shot stock-movement prediction and find that off-the-shelf ChatGPT underperforms specialized multimodal models when asked to forecast returns directly from tweets and prices, underscoring the trade-off between general reasoning and task-specific fine-tuning.

Against this backdrop, we adopt FinBERT as our core article-level sentiment engine. It is trained on financial language, provides probabilistic scores for positive, negative, and neutral tone, and is open-source and widely used in practice, which helps both performance and reproducibility in our setting.

## 2.3 LLM-Based Summarization and RAG in Finance

Summarization with large language models has become a central tool for compressing growing amounts of textual information into forms that are usable by both humans and quantitative models. In financial settings, LLMs are increasingly used to summarize earnings calls, annual reports, and news streams, with an emphasis on preserving key fundamentals, risk factors, and management guidance rather than simply reproducing headlines.

Recent applications often pair LLMs with retrieval-augmented generation (RAG). A retriever first selects the subset of documents relevant to a query, portfolio, or event; the generator then produces summaries or answers grounded in those sources. Compared with naïve summarization, RAG reduces hallucinations and encourages explicit linkage back to the underlying text, which is particularly important when working with legally or financially material disclosures. In parallel, studies such as Lopez-Lira and Tang (2023) and Xie et al. (2023) show both the promise and the limits of using LLMs directly for market prediction, reinforcing the value of architectures that tightly couple language models to clearly defined information sets.

Our news summary agent falls into this emerging RAG ecosystem. It uses embeddings to select a focused set of relevant articles, applies an LLM to transform those articles into a concise, event-focused narrative, and conditions the narrative on pre-computed sentiment scores from FinBERT. The goal is not only to compress text, but to highlight how news and sentiment interact with existing positions and risk exposures in a way that is transparent, auditable, and suitable for portfolio decision support.

# 3 Data and Sample Construction

## 3.1 Universe and Time Horizon

The initial investment universe consists of S&P 500 constituents. This choice reflects a balance between:

- rich and continuous news coverage,
- high liquidity and economic relevance, and
- manageable scope for a prototype dashboard.

The time horizon is driven by the historical coverage of the news APIs used. Conceptually, we view sentiment at a daily frequency and study behavior over a one-year window for visualization, with potential extension to multi-year backtests.

## 3.2 Data Sources

The system relies on three primary data sources:

**News data (Finnhub).** Finnhub provides historical and real-time news by ticker, including fields such as headline, summary, source, timestamp, and URL. This forms the backbone of our textual dataset.

**Market data and headlines (Yahoo Finance).** We use Yahoo Finance (via the `yfinance` interface) to gather daily open–high–low–close prices and volume for each stock, as well as supplemental headline feeds. This ensures that price and sentiment series are aligned on trading days.

**Regulatory filings (SEC).** SEC filings (10-K, 10-Q, etc.) are recognized as high-value textual sources. In the current prototype, they inform model design conceptually but are not yet integrated into the daily sentiment pipeline. They are a key component of our future work on a fundamental-analysis RAG agent.

## 3.3 Data Structure

We organize data around ticker–date pairs with three main tables:

- **Price table**: ticker, date, open, high, low, close, adjusted close, volume.
- **Article-level news table**: ticker, article ID, timestamp, source, headline/snippet, URL, and FinBERT outputs
$$P_i^{(+)},\ P_i^{(-)},\ P_i^{(0)},\ S_i.$$
- **Daily sentiment table**: ticker, date, aggregated daily sentiment $S_{j,t}^{\text{day}}$, 7-day rolling average $\bar{S}_{j,t}$, percentile ranks, and regime labels (Buy/Hold/Sell).

This relational structure supports both time-series analysis (per ticker) and cross-sectional views, and it extends naturally to portfolio-level aggregation.

### 3.4 Cleaning and Preprocessing

Key preprocessing steps include:

- **Ticker mapping**: enforce consistent ticker symbols and handle corporate actions where relevant.
- **Time alignment**: convert timestamps to a single timezone and map intraday news to trading dates (e.g., assigning late-evening headlines to the next trading day, if desired).
- **Deduplication**: remove duplicate or syndicated articles with nearly identical text or overlapping metadata.
- **Filtering**: optionally exclude extremely short or non-informative headlines.

## 4 Methodology and Models

### 4.1 Article-Level Sentiment with FinBERT

For each headline or short article $i$, we run FinBERT to obtain class probabilities:

$$P_i^{(+)} = P(\text{positive}), \quad P_i^{(-)} = P(\text{negative}), \quad P_i^{(0)} = P(\text{neutral}).$$

We convert these into a scalar sentiment score:

$$S_i = P_i^{(+)} - P_i^{(-)}, \qquad S_i \in [-1, 1]. \tag{1}$$

Interpretation:

- $S_i \approx +1$: strongly positive tone (e.g., large earnings beat, favorable guidance).
- $S_i \approx -1$: strongly negative tone (e.g., profit warning, regulatory penalty).
- $S_i \approx 0$: neutral or mixed tone.

To account for the strength of the model's prediction, we define a confidence weight:

$$w_i = \max\{P_i^{(+)}, P_i^{(-)}, P_i^{(0)}\} - \tfrac{1}{3}, \tag{2}$$

which measures how far the maximum class probability is from the uniform baseline. Higher $w_i$ indicates a clearer classification.

### 4.2 Daily Sentiment Aggregation

Consider ticker $j$ on trading day $t$, with $N_{j,t}$ associated articles. We define the daily sentiment as a weighted average:

$$S_{j,t}^{\text{day}} = \frac{\sum_{i=1}^{N_{j,t}} w_i \, S_i}{\sum_{i=1}^{N_{j,t}} w_i + \varepsilon}, \tag{3}$$

where $\varepsilon$ is a small constant to avoid division by zero when there are few or no articles.

Variants of this aggregation can incorporate:

- **Time-decay**: down-weight older intraday articles relative to recent ones.
- **Source quality**: assign higher $w_i$ to primary or more reputable news sources.

In the current prototype, we emphasize the simple, interpretable specification in (3).

### 4.3 Smoothing and Regime Classification

Daily sentiment can be noisy, so we compute a 7-day rolling average:

$$\bar{S}_{j,t} = \frac{1}{7} \sum_{k=0}^{6} S_{j,t-k}^{\text{day}}, \tag{4}$$

defined for dates with at least seven prior observations.

We then map $\bar{S}_{j,t}$ into regimes based on its empirical distribution over a look-back window (e.g., past one year). Let $q_{0.1}$ and $q_{0.9}$ denote the 10th and 90th percentiles of $\bar{S}_{j,t}$ for stock $j$. We define the signal as:

- **Sell / Short** if $\bar{S}_{j,t} < q_{0.1}$,
- **Hold** if $q_{0.1} \leq \bar{S}_{j,t} \leq q_{0.9}$,
- **Buy / Long** if $\bar{S}_{j,t} > q_{0.9}$.

This percentile-based approach normalizes for stock-specific baseline sentiment levels and focuses attention on unusually positive or negative periods rather than absolute levels.

### 4.4 News Ranking and Recent Headlines Panel

To contextualize sentiment spikes, we construct a Recent Headlines panel as follows:

1. Retrieve all articles for ticker $j$ over the last $D$ days (e.g., $D = 15$).
2. Deduplicate articles based on text similarity or matching metadata.
3. Sort remaining articles by timestamp in descending order.
4. Display the latest $K$ headlines (e.g., $K = 10$), including source and timestamp.

This simple ranking gives an immediate explanation of "what just happened" when the sentiment or price moves sharply.

### 4.5 News Summary Agent

The News Summary Agent is an LLM-based component that condenses multiple articles into a concise narrative.

#### Step 1: Retrieval and Embedding

For ticker $j$, we call a function such as

```
load_recent_news_for_ticker(j, max_days=15, max_articles=30).
```

Each article's text or snippet is converted into a vector embedding using an LLM-based embedding model. These embeddings support:

- filtering and deduplication,
- clustering similar news, and
- selecting representative content for summarization.

### Step 2: Hierarchical Summarization

We adopt a map–reduce-style pipeline:

- *Map stage*: For each article (or small group of articles), a GPT-4o-mini style model produces a short bullet-point summary emphasizing the main event, sentiment direction, and financial impact.
- *Reduce stage*: A second LLM call combines these bullet points into a higher-level synthesis that:
  - identifies recurring themes (e.g., earnings, regulation, product launches),
  - distinguishes one-off shocks from ongoing narratives,
  - aligns the textual story with the current sentiment regime (e.g., why sentiment is in the top decile).

Prompts explicitly instruct the model to prioritize financially material information, avoid fabricating numbers, and use concise, analyst-like language.

### Step 3: Output Formatting

The final summary for each ticker is a short paragraph or 3–5 bullet points summarizing:

- key recent developments,
- overall tone (positive/mixed/negative), and
- notable risks or uncertainties.

This is displayed alongside the sentiment and price panels in the dashboard.

# 5 Empirical Results

**Key empirical findings.** Over the sample from 1 November 2024 to 31 October 2025 the sentiment-driven Top-$K$ strategy compounds an initial investment of \$1 to approximately \$3.80, corresponding to a cumulative return of about 280% and an annualized return of roughly 287%. Over the same period SPY delivers a cumulative return of about 21% (annualized 21%). Despite exhibiting slightly higher volatility than SPY (23.3% vs. 19.7%), the sentiment strategy achieves a Sharpe ratio of 5.77 compared to 0.87 for SPY, and its maximum drawdown (–9.7%) is materially smaller than that of the benchmark (–18.8%). The strategy also records positive daily returns on 65.9% of trading days (versus 57.8% for SPY), indicating that the outperformance is not driven solely by a few extreme observations but by consistently better day-to-day outcomes.

## 5.1 Descriptive Statistics of Sentiment

We first examine the distribution of daily sentiment scores $S_{j,t}^{\text{day}}$ and rolling averages $\bar{S}_{j,t}$ across the sample.

- The cross-sectional distribution of $S_{j,t}^{\text{day}}$ has a mildly positive mean (about 0.06) and substantial dispersion (standard deviation around 0.36), with a 5th–95th percentile range of roughly $[-0.64, 0.84]$. This is consistent with the idea that most days feature neutral to mildly positive news, while a non-trivial fraction of days exhibit strongly positive or negative tone.

- The rolling averages $\bar{S}_{j,t}$ over a 7-day window preserve this positive bias but are much smoother (standard deviation around 0.15, with a 5th–95th range of approximately $[-0.18, 0.30]$), reflecting the aggregation of information over time and the tendency for sentiment to cluster in moderately positive or negative regimes.

Table 1 reports summary statistics for daily sentiment and its 7-day average.

Table 1: Summary Statistics of Sentiment

| Variable | Mean | Std. Dev. | 5th pct. | 95th pct. |
|---|---|---|---|---|
| $S_{j,t}^{\text{day}}$ | 0.0616 | 0.3552 | -0.6361 | 0.8425 |
| $\bar{S}_{j,t}$ (7-day) | 0.0619 | 0.1460 | -0.1803 | 0.3006 |

## 5.2 Sentiment and Price Dynamics: Case Study

To illustrate the interaction between sentiment and prices, we consider a representative large-cap stock such as Apple (AAPL) over a one-year window. Figure 6 plots the closing price together with daily sentiment bars and the 7-day rolling sentiment.

Figure 1: Illustrative Example: AAPL Price and Sentiment. Note: replace this placeholder with the actual figure from your system.

Qualitatively, we observe several recurring patterns:

- **Earnings events**: Around earnings announcements, sentiment typically spikes, with accompanying jumps or breaks in the price series. Positive surprises generate clusters of positive sentiment bars; negative surprises yield strong negative spikes.
- **Regulatory or legal news**: Regulatory scrutiny, lawsuits, or antitrust concerns often produce persistent negative sentiment, even if prices partially recover after the initial shock. This may reflect lingering uncertainty.
- **Quiet periods**: When there are few or no material headlines, sentiment hovers around zero, and the rolling average falls into the Hold regime, indicating an informationally quiet environment.

## 5.3 Same-Day Sentiment and Returns

We now move from descriptive behavior and strategy backtests to a more formal test of how our daily news sentiment relates to same-day stock returns. For each stock $j$ and trading day $t$, let

$$r_{j,t} = \log\left(\frac{P_{j,t}}{P_{j,t-1}}\right)$$

denote the close-to-close log return, and let $S_{j,t}^{\text{day}}$ be the aggregated FinBERT-based sentiment score for that ticker and date (implemented in the data as `score_mean`). We also track the total number of news items $N_{j,t}$ (variable `n_total`) as a measure of news-flow intensity.

### Single-Stock Time-Series Regressions

We begin with a representative S&P 500 constituent for which we have 266 trading days of matched price and sentiment data. We estimate the same-day predictive regression

$$r_t \;=\; \alpha \;+\; \beta\, S_t^{\text{day}} \;+\; \gamma\, N_t \;+\; \varepsilon_t, \tag{5}$$

using ordinary least squares with heteroskedasticity and autocorrelation robust (HAC / Newey–West) standard errors.

The OLS regression explains about 14.5% of the variation in daily returns ($R^2 = 0.145$). The coefficient on sentiment is positive and statistically highly significant:

$$\hat{\beta} = 0.0465, \qquad z = 3.92, \qquad p < 0.001.$$

Economically, this implies that moving the daily sentiment score up by one unit—for example from clearly negative to clearly positive news—is associated with an increase in same-day log return of roughly 4–5 basis points. More modest changes in sentiment still have non-trivial effects: a 0.2 increase in $S_t^{\text{day}}$ (e.g., from mildly negative to mildly positive) corresponds to about a 0.9 basis point shift in the expected same-day return.

By contrast, the coefficient on the news-count variable $N_t$ is economically very small and statistically insignificant:

$$\hat{\gamma} = 2.66 \times 10^{-5}, \qquad z = 0.39, \qquad p = 0.70,$$

indicating that *conditional on the tone of the news*, the raw number of articles has little incremental explanatory power for same-day returns in this specification.

Residual diagnostics indicate that returns are far from normal (strong skewness and excess kurtosis), and Ljung–Box tests detect some residual autocorrelation at short lags. This motivates two robustness checks: a robust regression that downweights outliers, and quantile regressions that look beyond the conditional mean.

**Robust Regression.** We re-estimate equation (5) using a Huber-type robust linear model (RLM). The robust fit yields

$$\hat{\beta}^{\text{RLM}} = 0.0387, \qquad z = 7.04, \qquad p < 0.001,$$

with an intercept close to zero and an insignificant coefficient on $N_t$. Both the sign and the magnitude of the sentiment effect are very similar to the OLS estimate, and the $z$-statistic is even larger. This suggests that the positive link between sentiment and returns is *not* driven by a small number of extreme return days or outlier news events.

**Distributional Effects via Quantile Regression.** To understand how sentiment affects different parts of the return distribution, we estimate quantile regressions for the 10th, 50th, and 90th percentiles:

$$Q_{r_t}(\tau \mid S_t^{\text{day}}, N_t) = \alpha(\tau) + \beta(\tau)\, S_t^{\text{day}} + \gamma(\tau)\, N_t, \qquad \tau \in \{0.1, 0.5, 0.9\}.$$

Across all three quantiles, the coefficient on sentiment remains positive and statistically significant:

$$\begin{aligned}
\tau = 0.10: &\quad \hat{\beta}(0.10) = 0.0642, \quad t = 5.53, \quad p < 0.001,\\
\tau = 0.50: &\quad \hat{\beta}(0.50) = 0.0294, \quad t = 4.81, \quad p < 0.001,\\
\tau = 0.90: &\quad \hat{\beta}(0.90) = 0.0413, \quad t = 3.22, \quad p = 0.001.
\end{aligned}$$

The effect is largest in the left tail ($\tau = 0.10$), suggesting that sentiment is particularly informative on poor-performance days: when news tone is more positive, the downside of returns is less severe. At the median and in the upper tail, the coefficients are smaller but still economically and statistically meaningful. In all three cases, the coefficient on $N_t$ remains close to zero and insignificant, reinforcing the conclusion that it is the *tone*, rather than the sheer volume, of news that drives the signal.

Overall, the single-stock evidence shows a robust, positive relationship between same-day sentiment and returns, affecting not only the conditional mean but the entire conditional distribution.

### Pooled Panel Regression with Ticker Fixed Effects

To exploit both cross-sectional and time-series variation across the S&P 500 universe, we estimate a pooled panel model with ticker fixed effects. The specification is

$$r_{i,t} \;=\; \alpha \;+\; \beta\, S^{\text{day}}_{i,t} \;+\; \gamma\, N_{i,t} \;+\; \delta_i \;+\; \varepsilon_{i,t}, \tag{6}$$

where $i$ indexes stocks, $\delta_i$ are stock-specific intercepts (fixed effects), and $\varepsilon_{i,t}$ are residuals. We use 132,965 daily observations with both price and sentiment available, and we report HAC standard errors with five lags.

In this panel setting, the estimated sentiment coefficient is again positive and highly statistically significant:

$$\hat{\beta}^{\text{panel}} = 0.0122, \qquad z = 74.22, \qquad p < 0.001.$$

The magnitude is smaller than in the single-stock case because $\hat{\beta}^{\text{panel}}$ represents a *common* within-stock sensitivity averaged across a large universe with heterogeneous volatility and news exposure. Interpreted economically, a 0.2 increase in the daily sentiment score for a given stock is associated with an increase in its same-day log return of about

$$0.2 \times 0.0122 \approx 0.0024,$$

or roughly 24 basis points on that day. While this is small relative to the overall volatility of daily equity returns, it is economically meaningful, and—as the strategy backtest in Section 5.5 illustrates—such differences can cumulate into substantial performance gaps when exploited systematically.

The coefficient on the news-count variable is slightly negative and statistically significant:

$$\hat{\gamma}^{\text{panel}} = -7.32 \times 10^{-5}, \qquad z = -2.78, \qquad p = 0.005,$$

but the effect is an order of magnitude smaller than that of sentiment and is not robustly economically large. A possible interpretation is that, after controlling for tone, days with very heavy news flow are often associated with elevated uncertainty or disagreement, which slightly depresses returns. However, the central message is that *sentiment tone* carries the dominant predictive content.

The panel regression attains an $R^2$ of about 4.9% (adjusted $R^2$ = 4.5%). This is in line with the broader return-prediction literature: daily returns are extremely noisy, so even variables with meaningful economic content typically explain only a few percent of the variance at this horizon. Against that backdrop, a highly significant, correctly signed $\hat{\beta}$ with this magnitude is a strong indication that the sentiment measure is picking up real information about same-day return realizations.

### Interpretation and Link to Trading Strategies

Across three complementary methods—OLS with HAC errors, robust regression, and quantile regression for a single stock, and a large-scale panel regression with ticker fixed effects—we obtain a consistent result: the same-day FinBERT-based sentiment score $S_{i,t}^{\text{day}}$ is positively related to same-day stock returns. This relationship holds after controlling for the number of news items, survives outlier-robust estimation, and appears not only in the conditional mean but also in the lower, median, and upper parts of the return distribution.

These findings underpin the trading results in our sentiment-based Top-$K$ strategy. The portfolio-level outperformance documented in Section 5.5 is not the result of an opaque black-box rule, but rather reflects an underlying micro-level pattern: on average, stocks with more positive same-day news sentiment earn higher same-day returns than those with neutral or negative sentiment. In this sense, the regressions validate that our sentiment signal has genuine *return-predictive content*, rather than being a purely descriptive or cosmetic metric.

## 5.4 Regime Behavior and Return Patterns

Using the percentile-based regime classification described in Section 4, each day for each stock is labeled as Buy/Long, Hold, or Sell/Short based on the 7-day rolling sentiment $\bar{S}_{j,t}$. For each $(j,t)$ with a defined regime, we compute the *next-day* stock return $R_{j,t+1}$ and then summarize these forward returns by regime.

Table 2 reports mean next-day returns, return volatility, and hit rates (fraction of days with positive return) for each regime over the full sample. All return and volatility figures are shown in percentage points per day (so, for example, 0.02 corresponds to 0.02% per day).

Table 2: Next-Day Returns by Sentiment Regime

| Regime | Mean Return | Std. Dev. | Hit Rate (Return > 0) | # Obs. |
|---|---|---|---|---|
| Sell / Short ($\bar{S}_{j,t} < q_{0.1}$) | 0.02 | 2.41 | 51.84% | 13,014 |
| Hold ($q_{0.1} \leq \bar{S}_{j,t} \leq q_{0.9}$) | 0.02 | 2.19 | 50.96% | 103,323 |
| Buy / Long ($\bar{S}_{j,t} > q_{0.9}$) | -0.01 | 2.06 | 49.95% | 13,029 |

Two features stand out. First, there is no clear monotone pattern in which higher sentiment translates into higher next-day returns. If anything, the bottom-sentiment decile (Sell/Short) exhibits slightly higher average next-day returns and a marginally higher hit rate than the top-sentiment decile, while the Buy/Long regime has a slightly negative average next-day return and the lowest hit rate. Second, the differences in mean returns and hit rates across regimes are small relative to the very high day-to-day volatility of individual stocks (standard deviations on the order of 2–2.5 percentage points per day).

Overall, these results suggest that simple unconditional decile classifications of $\bar{S}_{j,t}$ have limited predictive power for next-day stock moves in our sample. This motivates the more structured portfolio-level experiments in the next subsection, which focus on cross-sectional ranking within a given day and on realistic trading constraints (lagged information and weekly rebalancing).

## 5.5 Portfolio Strategy and Backtesting

In this subsection we move from descriptive signal properties to explicit trading strategies. We start with an idealized daily rebalanced strategy that exploits same-day sentiment and then introduce two more realistic variants that (i) use only lagged information and (ii) rebalance weekly. All strategies are evaluated on daily returns and compared to a passive SPY benchmark.

### 5.5.1 Strategy Definitions and Implementation Constraints

**Daily Top-$K$ strategy (idealized).** The baseline strategy (used in earlier experimentation) is a daily sentiment–sorted portfolio. For each asset $i$ on day $t$ we compute an average sentiment score $s_{i,t}$ from FinBERT-scored news headlines (e.g. Finnhub and Yahoo Finance):

$$s_{i,t} = \text{score\_mean}_{i,t}.$$

On each trading day $t$ we rank the cross-section in descending order of $s_{i,t}$ and select the top $K$ names. In the baseline configuration we set $K = 3$. The portfolio weights are

$$w_{i,t} = \begin{cases} \frac{1}{K}, & \text{if asset } i \text{ is among the top } K \text{ sentiment names on day } t, \\ 0, & \text{otherwise.} \end{cases}$$

Daily close-to-close returns are

$$r_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}},$$

and the portfolio return is

$$R_{p,t} = \sum_i w_{i,t}\, r_{i,t}.$$

Starting from $V_0 = 1$, the equity curve evolves as

$$V_t = V_{t-1}\big(1 + R_{p,t}\big), \qquad t = 1, \ldots, T.$$

Because this strategy trades every day on same-day sentiment, it should be viewed as an *upper bound* on the economic value of the signal rather than a realistic implementable strategy.

**Weekly Top-$K$ strategy with previous-day sentiment.** To incorporate more realistic frictions, we first impose weekly rebalancing and lagged information. We define a set of rebalance dates $\mathcal{T} = \{t_1, \ldots, t_M\}$ as the first trading day of each calendar week in the sample. For each rebalance date $t \in \mathcal{T}$ we construct a signal based on *previous-day* sentiment:

$$s^{\text{lag}}_{i,t} = s_{i,t-1},$$

where $t - 1$ is the last trading day of the previous week (typically Friday). We then rank assets by $s^{\text{lag}}_{i,t}$ and form an equal-weight portfolio of the top $K$ names:

$$w^{\text{wk}}_{i,t} = \begin{cases} \frac{1}{K}, & \text{if } i \text{ is in the top } K \text{ by } s^{\text{lag}}_{i,t}, \\ 0, & \text{otherwise.} \end{cases}$$

These weights are held constant until the next rebalance date, so that the daily portfolio return between two rebalance dates $t$ and $t'$ is

$$R_{p,\tau} = \sum_i w^{\text{wk}}_{i,t}\, r_{i,\tau}, \qquad \tau = t + 1, \ldots, t'.$$

We sweep over $K \in \{1, \ldots, 10\}$ and report the best-performing configuration (in terms of Sharpe ratio), which occurs at $K = 10$.

**Weekly Top-$K$ strategy with recency-weighted weekly sentiment.** The previous-day strategy still depends on a single headline day. As a more "fundamental-style" signal we also aggregate sentiment across the entire previous week. Let $t$ be a rebalance date and let $\{t-5,\dots,t-1\}$ denote the previous week's trading days (Monday to Friday). For each stock $i$ we compute a recency-weighted average of daily sentiment,

$$\tilde{s}_{i,t}^{\text{week}} = \sum_{k=1}^{5} \omega_k \, s_{i,t-k}, \qquad \sum_{k=1}^{5} \omega_k = 1, \quad \omega_5 > \cdots > \omega_1, \tag{7}$$

where the weights $\omega_k$ increase with recency so that Friday carries the highest weight and Monday the lowest. In the implementation we use simple linear weights,

$$\omega_k = \frac{k}{1+2+3+4+5}, \qquad k = 1,\dots,5.$$

We then rank assets by $\tilde{s}_{i,t}^{\text{week}}$ and form an equal-weight portfolio of the top $K$ names at each rebalance date $t$, again holding weights constant until the next week. A grid search over $K \in \{1,\dots,10\}$ identifies $K = 9$ as the best trade-off between return and risk for this weekly-averaged signal.

#### 5.5.2 Benchmark

As a market benchmark we use the SPDR S&P 500 ETF (SPY). Daily SPY prices are obtained from Yahoo Finance. Let $P_t^{\text{SPY}}$ be the adjusted closing price on day $t$; daily returns are

$$R_{\text{SPY},t} = \frac{P_t^{\text{SPY}} - P_{t-1}^{\text{SPY}}}{P_{t-1}^{\text{SPY}}}.$$

For visualization we normalize all equity curves—the daily sentiment strategy, the two weekly strategies, and SPY—to start at 1 on the first common date.

#### 5.5.3 Performance Metrics

Let $\{R_t\}_{t=1}^{T}$ denote a series of daily portfolio returns and assume 252 trading days per year. We report the following metrics:

**Cumulative and annualized return.**

$$\text{CumRet} = \prod_{t=1}^{T} (1+R_t) - 1, \qquad \text{AnnRet} = (1+\text{CumRet})^{252/T} - 1.$$

**Annualized volatility.** With $\sigma_{\text{daily}}$ the standard deviation of daily returns,

$$\text{AnnVol} = \sigma_{\text{daily}} \sqrt{252}.$$

**Sharpe ratio.** Given an annual risk-free rate $r_f$ (we use $r_f = 4\%$), the implied daily risk-free rate is

$$r_{f,\text{daily}} = (1+r_f)^{1/252} - 1.$$

Excess returns are $R_t^{\text{ex}} = R_t - r_{f,\text{daily}}$, and the annualized Sharpe ratio is

$$\text{Sharpe} = \frac{\mathbb{E}[R_t^{\text{ex}}]}{\sqrt{\text{Var}(R_t^{\text{ex}})}} \sqrt{252}.$$

**Maximum drawdown.** Let $V_t = \prod_{\tau \le t}(1 + R_\tau)$ be the wealth process and $P_t = \max_{\tau \le t} V_\tau$ its running peak. The drawdown at $t$ is

$$\mathrm{DD}_t = \frac{V_t}{P_t} - 1,$$

and the maximum drawdown is $\mathrm{MaxDD} = \min_t \mathrm{DD}_t$.

**Hit rate.** The hit rate (win percentage) is

$$\mathrm{HitRate} = \frac{\#\{t : R_t > 0\}}{T}.$$

### 5.5.4 Summary Statistics and Comparison with SPY

Table 3 reports performance metrics for the three sentiment strategies and SPY over the sample from 1 November 2024 to 31 October 2025. The *Daily Top-3* strategy uses idealized same-day information and daily rebalancing, while the two weekly variants use only lagged information and rebalance at the start of each week.

Table 3: Performance metrics: daily and weekly sentiment strategies vs. SPY

| Strategy | Cum. Return | Ann. Return | Ann. Vol | Sharpe | Max DD | Hit Rate | # Days |
|---|---|---|---|---|---|---|---|
| Daily Top-3 sentiment | 2.8037 | 2.8655 | 0.2326 | 5.7727 | -0.0971 | 0.6586 | 249 |
| Weekly prev-day Top-10 | 0.1757 | 0.1780 | 0.1807 | 0.7793 | -0.1872 | 0.5181 | 249 |
| Weekly weekavg Top-9 | 0.1124 | 0.1134 | 0.1802 | 0.4676 | -0.1705 | 0.5160 | 250 |
| SPY | 0.2089 | 0.2117 | 0.1966 | 0.8744 | -0.1876 | 0.5783 | 249 |

The results highlight three points. First, the daily Top-3 strategy delivers very high risk-adjusted performance (Sharpe $\approx 5.8$) and triples capital over the sample, but at the cost of extreme turnover and tight timing assumptions. Second, once we restrict ourselves to weekly rebalancing and lagged information, the alpha largely dissipates: the Weekly prev-day Top-10 strategy achieves a Sharpe ratio of 0.78, while the Weekly weekavg Top-9 strategy is weaker still (Sharpe 0.47). Finally, the weekly strategies are broadly comparable to, or slightly worse than, a passive SPY benchmark in both total return and Sharpe, underscoring how implementation realism erodes the “paper” alpha of the raw sentiment signal.

### 5.5.5 Visualizing Performance and the $K$-Grid Search

Figure 2 plots the equity curve of the Daily Top-3 strategy against SPY. Figures 3 and 4 show the two weekly strategies versus SPY, with all series rebased to 1 at the start of the sample.

To understand the role of portfolio concentration, we perform a simple grid search over $K \in \{1, \ldots, 10\}$ for each weekly strategy. For each $K$ we recompute the full time series of daily returns and the performance metrics above. Figure 5 plots the Sharpe ratio as a function of $K$. For the Weekly prev-day strategy, Sharpe peaks around $K = 10$, while the Weekly weekavg strategy attains its best Sharpe around $K = 9$. Very small $K$ values generate more volatile portfolios with lower Sharpe ratios, while very large $K$ values dilute the signal and compress both return and risk.

Overall, these experiments suggest that the FinBERT-based sentiment signal is economically strong at short horizons, but its predictive content decays quickly and is sensitive to realistic trading constraints.

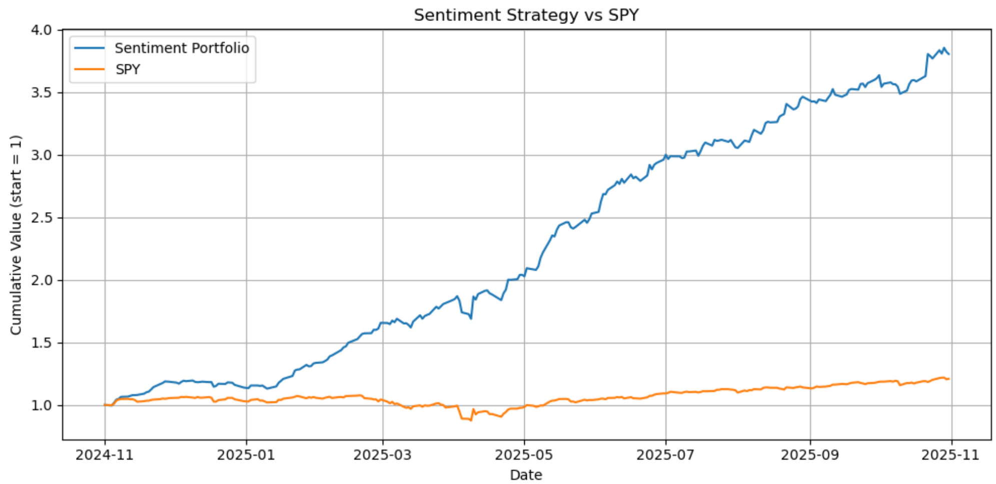


Figure 2: Daily Top-3 sentiment strategy vs. SPY. All equity curves are rebased to 1 at the start of the sample.

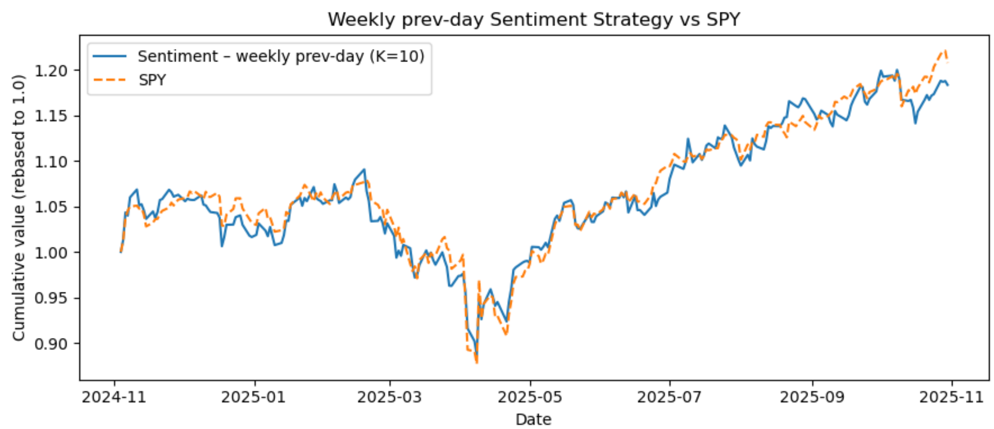


Figure 3: Weekly prev-day Top-10 sentiment strategy vs. SPY. Portfolio weights are updated weekly using previous-day sentiment.

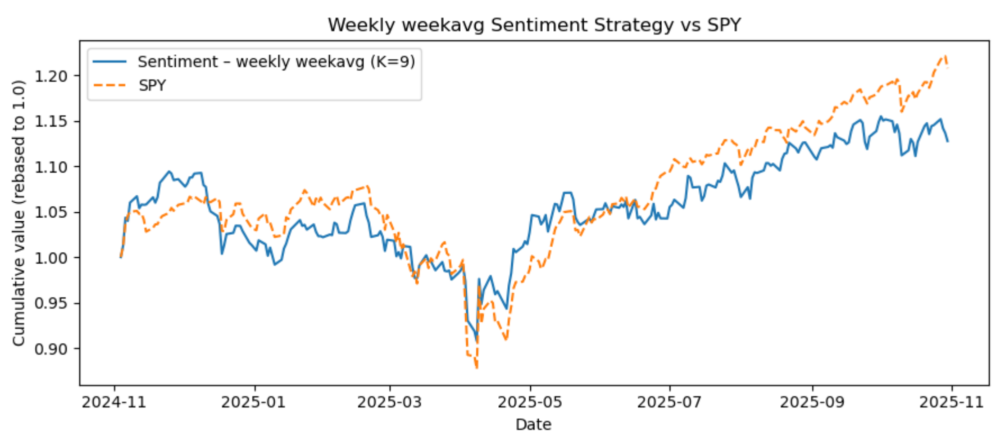


Figure 4: Weekly week-averaged Top-9 sentiment strategy vs. SPY. Signals use a recency-weighted average of the previous week's daily sentiment.

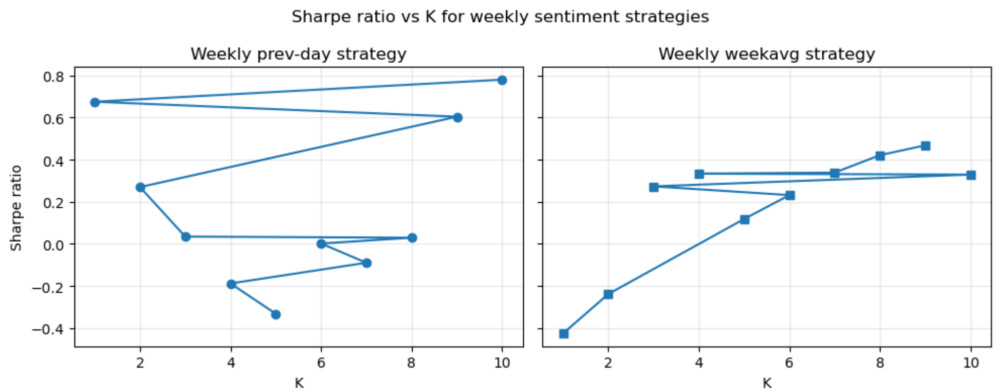


Figure 5: Sharpe ratio as a function of $K$ for the weekly sentiment strategies. The Weekly prev-day strategy attains its maximal Sharpe at $K = 10$, while the Weekly weekavg strategy peaks at $K = 9$.

### 5.6 Sentiment and Price Dynamics: Case Study

To illustrate the interaction between sentiment and prices, we consider a representative large-cap stock such as Apple (AAPL) over a one-year window. Figure 6 plots the closing price together with daily sentiment bars and the 7-day rolling sentiment.

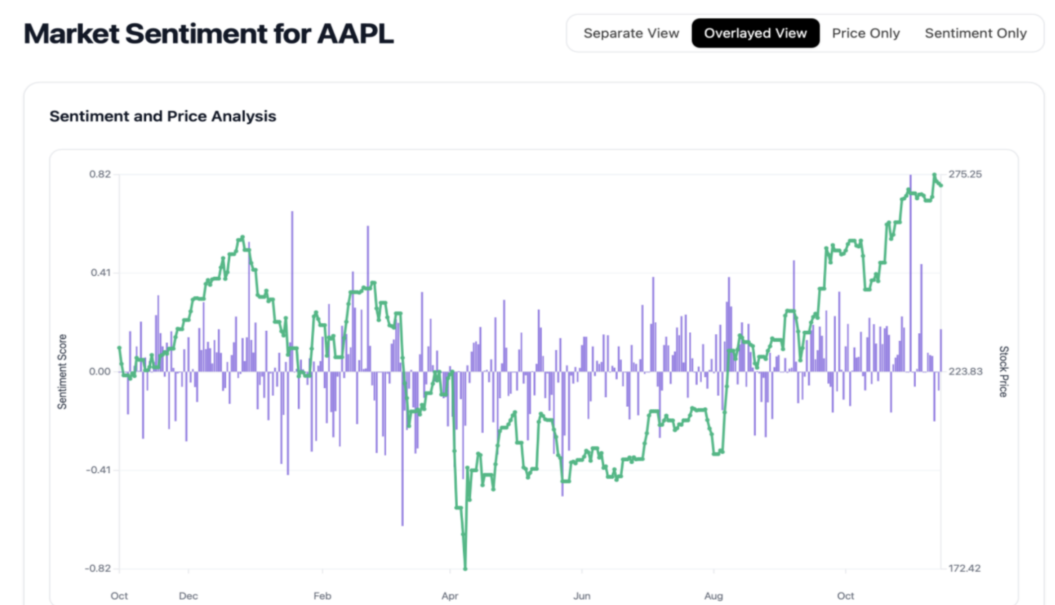


Figure 6: Illustrative Example: AAPL Price and Sentiment.

Qualitatively, we observe several recurring patterns:

- **Earnings events**: Around earnings announcements, sentiment typically spikes, with accompanying jumps or breaks in the price series. Positive surprises generate clusters of positive sentiment bars; negative surprises yield strong negative spikes.
- **Regulatory or legal news**: Regulatory scrutiny, lawsuits, or antitrust concerns often produce persistent negative sentiment, even if prices partially recover after the initial shock. This may reflect lingering uncertainty.

- **Quiet periods**: When there are few or no material headlines, sentiment hovers around zero, and the rolling average falls into the Hold regime, indicating an informationally quiet environment.

### 5.7 User-Facing Insights from the Dashboard

Beyond numerical performance, the platform delivers several practical benefits to users:

- A one-year sentiment–price panel that provides fast visual diagnostics on how market tone has evolved around a stock.
- A Live Market Insights panel that converts sentiment into simple, intuitive signals (Buy/Hold/Sell) without forcing users to interpret raw scores.
- A news summary agent that turns dozens of articles into a short narrative, saving analyst time and reducing cognitive load.

In practice, these components can be used together for tasks such as monitoring watchlists, diagnosing large price moves, and screening for names with unusually positive or negative recent news.

## 6 Interactive Dashboard and Front-End Architecture

### 6.1 Overview

To translate the sentiment pipeline into a practical analytical tool, we developed an interactive web-based dashboard that integrates all components of the system—including sentiment signals, recent headlines, and LLM-generated summaries—into a unified interface. The dashboard is implemented in Next.js and React, emphasizing fast navigation, concise information flow, and consistent interpretation of model outputs. Its purpose is not only to visualize results but to support rapid diagnostic assessment of the news environment surrounding any stock in the S&P 500 universe.

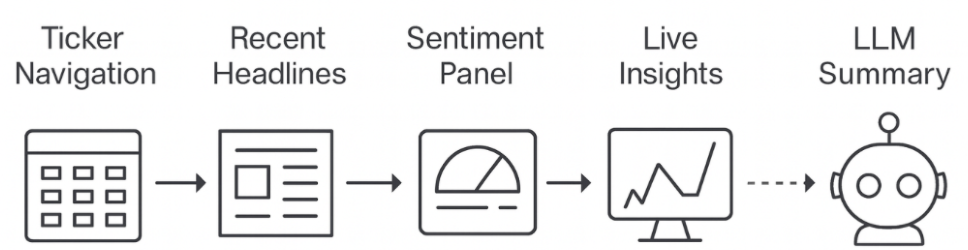


Figure 7: Components of the front-end

### 6.2 Unified Ticker Navigation

The user experience begins with a ticker-navigation page that displays all S&P 500 constituents in a grid layout, accompanied by a search bar for direct lookup. Selecting a ticker triggers a real-time load of its sentiment series, news panel, and summary outputs. This design mirrors the workflow of financial data terminals while remaining lightweight and responsive. It allows quick switching across names, enabling portfolio managers or analysts to move seamlessly from one stock's information environment to another.

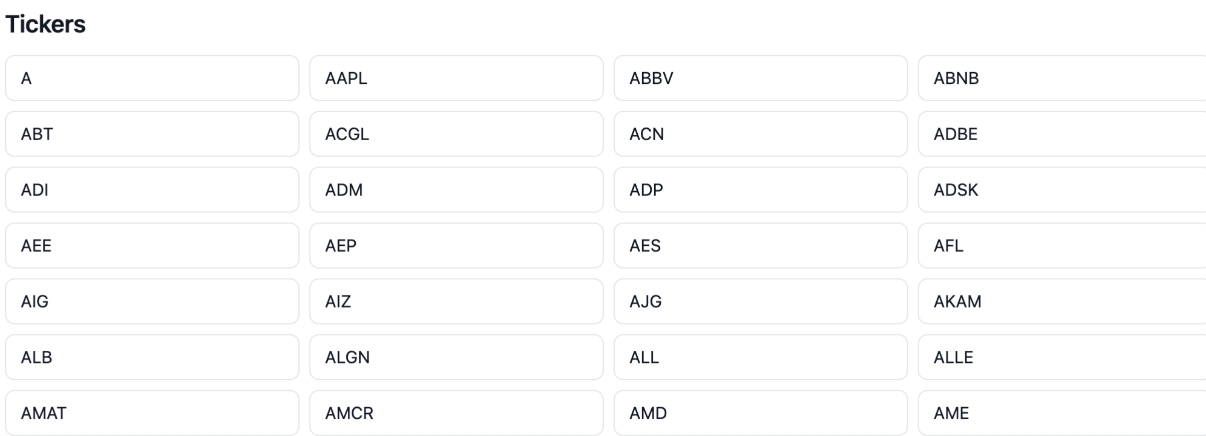


Figure 8: Interactive Ticker Navigation Grid - The dashboard homepage displays an interactive grid of S&P 500 tickers, enabling users to quickly browse, search, and select any stock. Choosing a ticker initiates the full data-loading workflow

## 6.3 Stock Sentiment Panel

Once a ticker is selected, the system presents a sentiment panel that combines recent headlines with model-generated sentiment and price information. Daily sentiment values computed from FinBERT-based article scores are visualized, while the stock's closing price is plotted on a secondary axis. The example is of one year horizon. The interface supports multiple viewing modes, including price-only, sentiment-only, and an overlaid view. This flexibility helps users evaluate whether sentiment changes coincide with price movements and whether deviations arise around major events such as earnings announcements, guidance revisions, or regulatory actions.

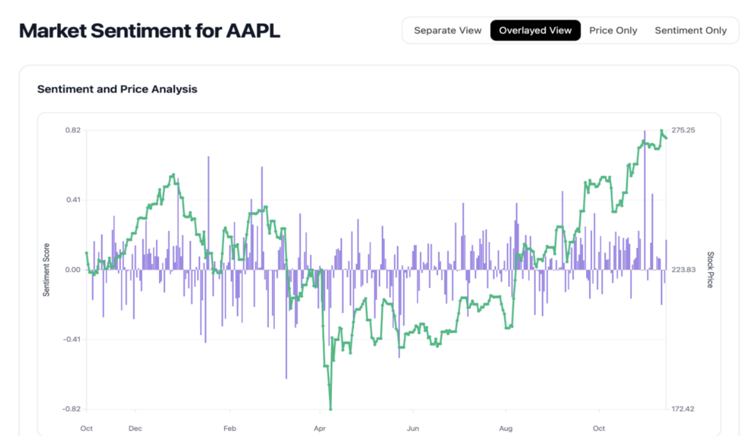


Figure 9: Stock Sentiment Panel - The sentiment–price panel for a selected ticker, combining daily FinBERT-based sentiment bars with the stock's closing price. The interface supports multiple viewing modes (sentiment-only, price-only, or combined) to highlight co-movements and event-driven deviations.

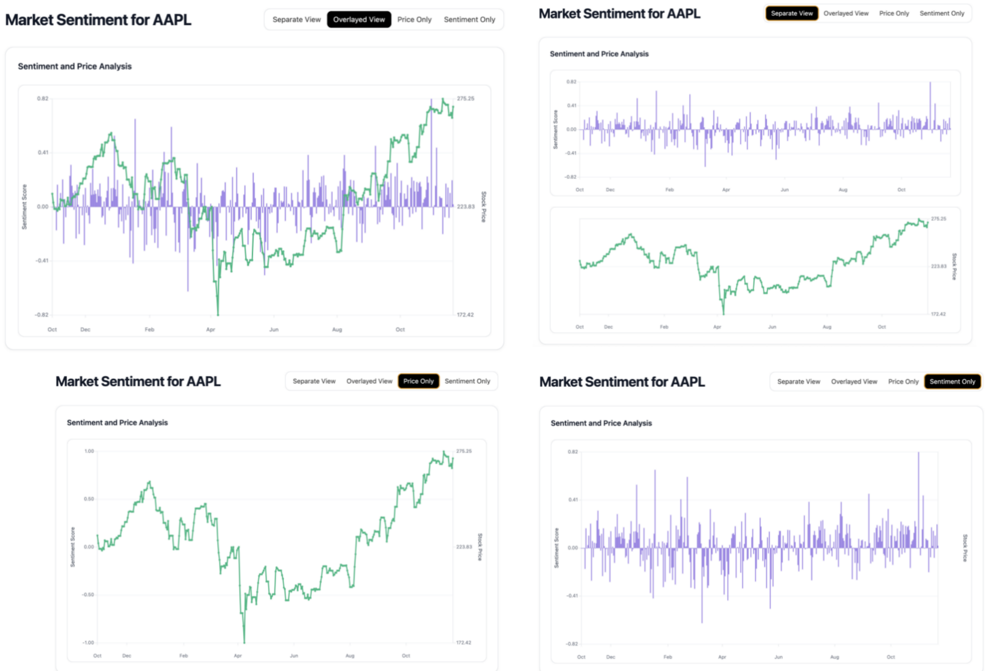


Figure 10: Multi-view: The overlayed view highlights co-movement patterns, while the separated charts isolate each signal for clearer inspection. The price-only and sentiment-only modes enable focused analysis of trend behavior and sentiment volatility, respectively.

### 6.4 Live Market Insights

The Live Market Insights panel provides a concise summary of the key metrics generated by the system. It reports the most recent daily sentiment score, the 7-day rolling sentiment average, an advisory classification (Buy/Hold/Sell) based on percentile thresholds, and the number of news items accumulated in the period. These metrics help contextualize the sentiment signal and communicate its current informational relevance at a glance. By summarizing the system's quantitative outputs into an easily interpretable format, this panel allows users to form a high-level assessment of market tone within seconds.

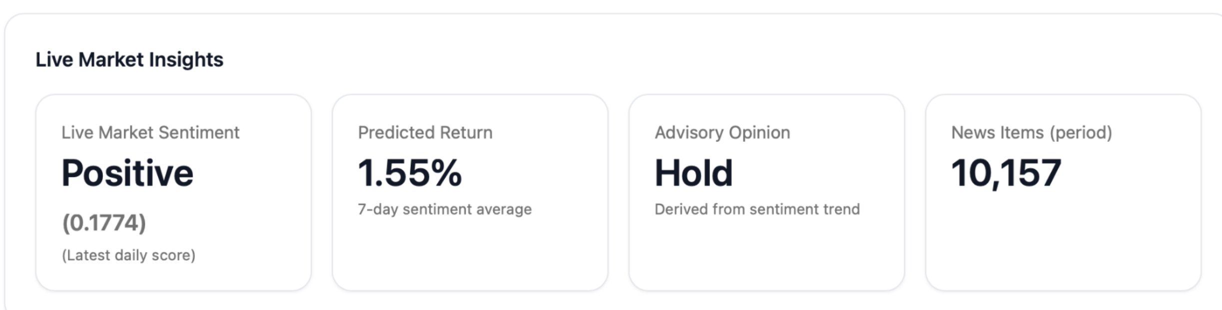


Figure 11: Live Market Insights Panel - The Live Market Insights summary, reporting the latest daily sentiment score, the 7-day rolling sentiment average, an advisory classification (Buy/Hold/Sell), and the total number of news items observed in the selected period.

### 6.5 Recent-headlines

To make the sentiment signal interpretable, the dashboard also includes a Recent Headlines module. For each ticker, the system retrieves articles from multiple sources over a short look-back window (e.g., 15 days), deduplicates overlapping entries, and sorts them by timestamp. Each headline is displayed alongside its publication date, source, and sentiment label, and clicking the source redirects the user to the original article. This component clarifies which specific events contributed to recent sentiment readings and can also serve as a streamlined news-reading tool within the dashboard.

**Recent Headlines for AAPL**

The table below gives each of the most recent headlines of the stock and the negative, neutral, positive aggregated sentiment score.

| Date | Headline | Source | Sentiment |
|---|---|---|---|
| 2025-11-15 | Apple Just Secured More Than Half of Taiwan Semi's Most Valuable Asset | 24/7 Wall St. | Positive (0.1948) |
| 2025-11-15 | Ex-Fed Governor Kugler Faced Ethics Probe Before Resignation | Bloomberg | Strong Negative (-0.4600) |
| 2025-11-15 | This Critical AI Stock Is Now Building Factories in the U.S. | 24/7 Wall St. | Positive (0.1420) |
| 2025-11-15 | BofA Reaffirms Bullish Stance on Alphabet (GOOGL) As AI And Cloud Growth Accelerate | Insider Monkey | Strong Positive (0.8083) |
| 2025-11-15 | Engie (ENXTPA:ENGI) Valuation: Assessing the Impact of Major Renewable Energy Agreements and Capacity Growth | Simply Wall St. | Positive (0.1192) |
| 2025-11-15 | Did Engie's (ENXTPA:ENGI) 15-Year Apple Renewable Deal Just Redefine Its Revenue Strategy? | Simply Wall St. | Positive (0.2472) |

Figure 12: Recent Headlines Panel - The Recent Headlines module displaying deduplicated, timestamp-ordered news items with source and sentiment labels. This component provides transparency into the specific events contributing to short-term sentiment shifts.

### 6.6 LLM News Summary Agent Output

To integrate qualitative context with quantitative sentiment, the dashboard provides a dedicated interface for the LLM-based summary agent. This module synthesizes the most relevant recent articles into a compact narrative that highlights key themes, tone, risks, catalysts, and regulatory developments. Rather than duplicating raw headlines, the summary expresses the central narrative driving recent sentiment and explains major movements in the sentiment or price signals. In practice, this substantially reduces the time analysts need to understand the news landscape surrounding a ticker, especially during periods of dense or rapidly evolving information flow.

In this example, the user selects AAPL from the homepage ticker grid, and the system immediately generates a structured research-style summary using news from the past 15 days. The output on the right demonstrates the quality of the LLM-generated summarization: it is concise, financially relevant, and organized around analyst-friendly categories such as key themes, tone, risks, catalysts, and regulatory considerations. Importantly, the summary does not merely repeat headlines—it synthesizes multiple articles into a coherent narrative. For AAPL, the model correctly identifies the overarching drivers (iPhone sales momentum, market expansion in the U.S. and China), distinguishes tone (bullish), highlights risk factors (macroeconomic headwinds, manufacturing exposure in Asia), and concludes with a reasonable final assessment. This level of structure and information compression mirrors how human analysts would produce a morning note, making it useful for quickly understanding the recent news flow behind a stock. Overall, the ex-

```
# LLM prompt
prompt = f"""
You are an equity research analyst.

You are given multiple recent news items about the ticker {ticker}
from the last {max_days} days.

Write a concise research-grade summary in **5–8 bullet points**:
- Key themes
- Tone (bullish / bearish / mixed)
- Risks & catalysts
- Regulator or guidance notes
- Final view (positive / neutral / negative)

Recent news:
{news_text}
"""
```

Figure 13: LLM News Summary Agent Prompt - The interface for the LLM-based news summary agent, which synthesizes multiple recent articles into a concise, analyst-oriented narrative highlighting key themes, tone, risks, catalysts, and regulatory considerations.

ample illustrates that the summarization agent consistently produces well-organized, readable, and contextually grounded commentary, enabling analysts to grasp the essential story within seconds rather than parsing dozens of individual articles.

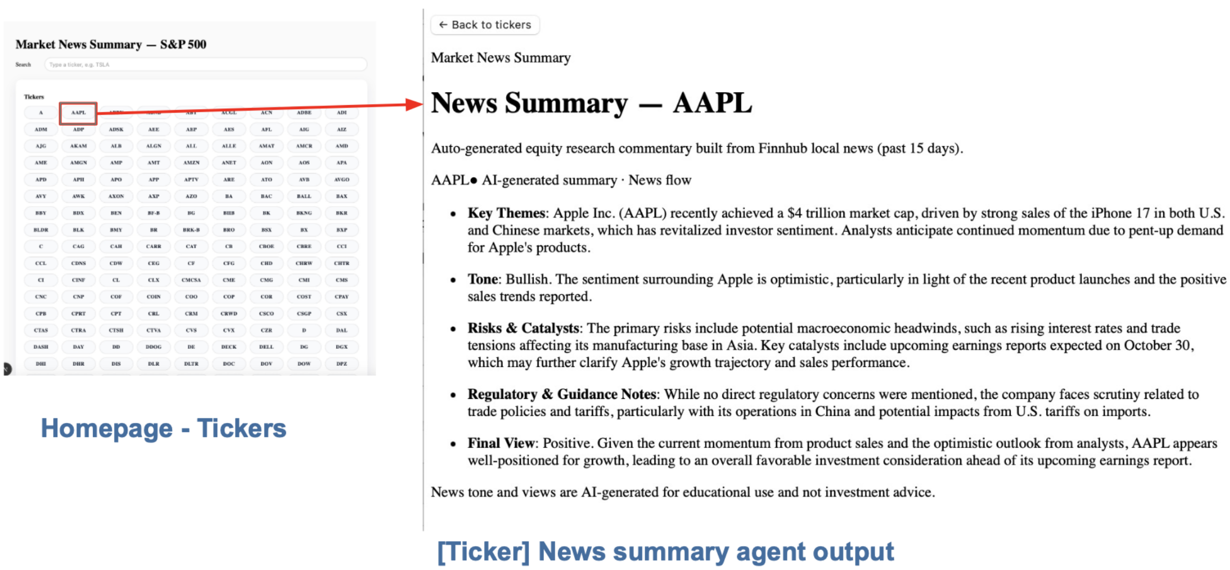


Figure 14: Example of the News Summary Agent Workflow: The user selects a ticker (AAPL) from the homepage grid, after which the system retrieves recent articles from the past 15 days and generates an AI-driven, analyst-style summary. The output synthesizes dispersed news into structured insights—including key themes, tone, risks, catalysts, and a final view—demonstrating the agent's ability to convert raw headlines into a coherent research narrative.

# 7 Limitations and Future Work

Despite its functionality, the current system has several limitations:

- **Limited backtesting**: We implement a simple long-only Top-$K$ sentiment strategy as a proof of concept, but do not yet conduct a full-scale evaluation with transaction costs, turnover constraints, risk limits, and alternative portfolio constructions.

- **Coverage heterogeneity**: Some firms and sectors generate abundant news, while others are sparsely covered, affecting comparability of sentiment intensity.
- **Model robustness**: Financial sentiment models can be vulnerable to adversarial phrasing. We do not yet explicitly guard against manipulation or adversarial news.
- **Summarization evaluation**: We have not yet quantified summary quality using standard NLP metrics or extensive human evaluation.
- **Limited integration of fundamentals**: SEC filings and structured financial data are not yet fully integrated into the live sentiment and summarization pipeline.

# 8 Directions for future work

Looking forward, several research directions can substantially broaden the analytical capabilities and institutional value of the platform. One natural extension is the evolution from a single summarization agent to a multi-agent research framework, where specialized LLM-powered components collaborate to process and interpret financial information. Inspired by the multi-modal agent architecture shown below, the system could incorporate dedicated agents for summarization, contextual question answering, data processing, valuation diagnostics, and portfolio monitoring. Such a modular design would allow each agent to specialize in a specific modality—unstructured text, structured financial data, audio transcripts, or regulatory filings—while exchanging information through a shared memory layer or retrieval-augmented system. This shift would transform the platform into a coordinated research assistant capable of interacting with heterogeneous data sources and performing complex analytic tasks beyond single-document summarization.

A second major direction involves retrieval-augmented analysis of fundamental information. While the current system focuses on news-based sentiment, a richer understanding of a firm's outlook requires integrating long-form disclosures such as 10-Ks, 10-Qs, earnings presentations, and cross-sectional financial statements. Building a RAG index over these materials would enable the LLM to ground its summaries and explanations in factual, domain-specific content. As a result, the platform could move toward generating outputs that resemble full equity research notes—linking recent news to valuation metrics, historical fundamentals, and structural trends within the firm's operating environment.

Another realm is earnings call integration. Earnings calls are among the most information-dense corporate disclosures, and incorporating both text transcripts and audio recordings would allow the platform to detect tonal cues, shifts in executive confidence, and changes in narrative emphasis across quarters. Aligning call analysis with news-driven sentiment would provide a unified view of event-specific narratives, particularly around guidance revisions, macro commentary, and competitive positioning.

Beyond firm-level analysis, future work should expand toward portfolio- and sector-level sentiment intelligence. Aggregating sentiment across holdings would allow portfolio managers to diagnose concentrated exposures to negative news or regulatory risk while identifying cross-sectional divergences in sentiment dynamics. Sector-level indices could support macro research and factor modeling, enabling the platform to serve not only as a tool for single-name monitoring but as a broader risk-management signal.

Methodologically, the sentiment model itself can be extended beyond a single scalar score. A multi-dimensional sentiment representation—capturing regulatory pressure, macroeconomic uncertainty, product-specific developments, competitive dynamics, or supply-chain disruptions—would

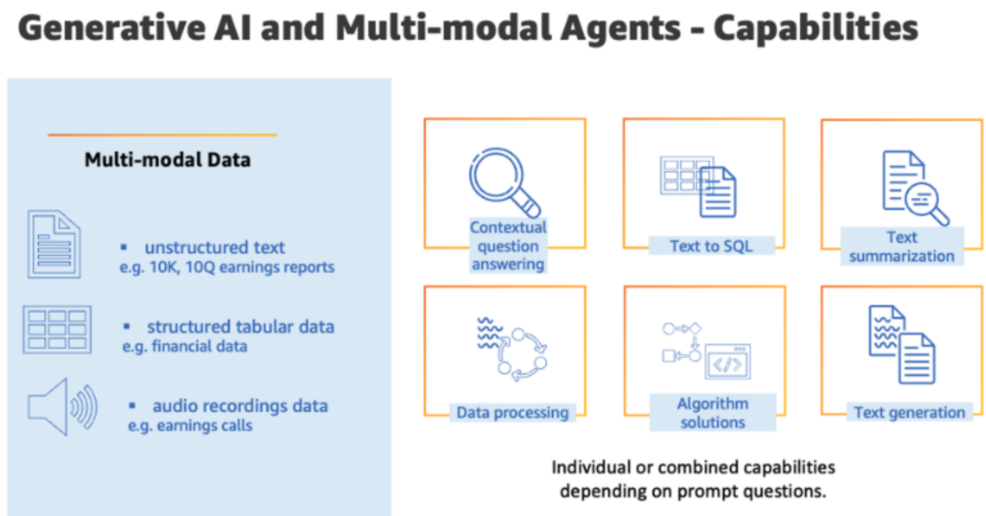


Figure 15: This figure, adapted from AWS (2023) Amazon Web Services' technical article on generative AI and multi-modal agents in financial markets, illustrates how modern AI systems can integrate heterogeneous data types—including unstructured text, structured tabular data, and audio transcripts—into coordinated multi-agent workflows. These agents support capabilities such as contextual question answering, text summarization, data processing, text-to-SQL translation, and algorithmic reasoning. The framework highlights the relevance of multi-modal, multi-agent architectures for future extensions of our platform, enabling richer analysis across diverse financial information sources.

allow more nuanced interpretations and support event-window attribution around major catalysts. Combining this with entity-level extraction and thematic clustering would provide deeper explanatory structure, offering analysts a clearer view of what precisely is driving sentiment shifts.

The system can also benefit from graph-based modeling of news and entities, drawing inspiration from pipelines such as the Neo4j Graph Data Science example. By constructing a graph that links companies, events, themes, and co-mentioned entities, the platform could identify central nodes of risk or influence, trace sentiment spillovers across supply chains or peer groups, and detect emerging clusters of market concern. Graph neural networks or graph-based ranking algorithms would allow more sophisticated inference than linear sentiment aggregation.

A complementary avenue for expansion involves integrating alternative datasets, including social media sentiment, customer reviews, job postings, shipping activity, and web traffic analytics. These sources, when denoised and weighted appropriately, can provide early signals ahead of traditional news flow. Their inclusion would further enhance the predictive and diagnostic capabilities of the system, especially in rapidly evolving sectors such as technology or consumer discretionary.

Finally, system-level improvements are essential for real-world deployment. Enhancing the platform with real-time streaming, GPU-accelerated inference, vector-database retrieval, and caching strategies would enable low-latency updates and scalable operation. Coupled with natural-language query interfaces—allowing users to ask questions such as "Why did sentiment for NVDA deteriorate last week?"—the dashboard could evolve into an interactive research environment that supports both rapid monitoring and deep investigation.

Taken together, these extensions outline a path toward a comprehensive, multi-agent, multi-modal financial intelligence platform. By combining retrieval-augmented reasoning, advanced sentiment modeling, event-level analysis, graph structures, and alternative data, future iterations of

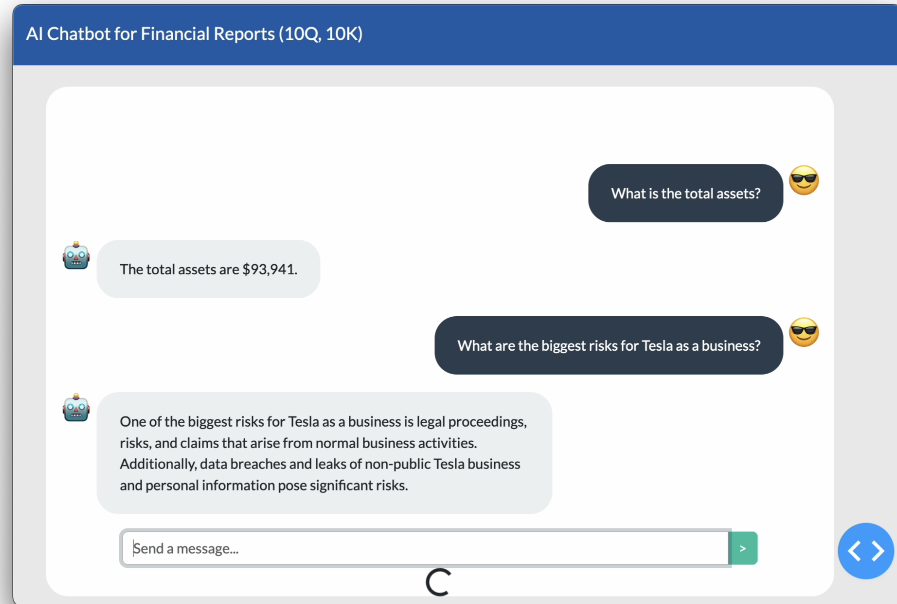


Figure 16: This figure, adapted from Kharshit (2024), illustrates how a retrieval-augmented generation system can answer detailed queries about financial disclosures such as 10-Q and 10-K filings. By retrieving relevant sections from financial statements and passing them to a large language model, the chatbot can provide grounded answers regarding total assets, business risks, and other firm-specific metrics.

the system can deliver richer insights, stronger explanatory power, and more actionable intelligence for analysts, portfolio managers, and risk teams.

# 9 Conclusion

We developed an end-to-end system for news-driven sentiment extraction, summarization, and visualization tailored to equity markets. By combining FinBERT-based article-level classification and a transparent aggregation procedure that converts textual information into daily sentiment indicators together with a large language model summary agent, the platform provides an interpretable layer on top of real-time news flow. The interactive dashboard allows investors to contextualize sentiment movements alongside historical price behavior and to rapidly understand narrative drivers as well as assess whether a stock's recent tone is unusual relative to its own history.

Beyond demonstrating feasibility, the project highlights how lightweight natural language processing components and modern large language model interfaces can be integrated into asset-management workflows while maintaining transparency. Although parts of our empirical analysis remain descriptive, the system also includes a portfolio-level component that transforms sentiment signals into actionable insights. This addition moves the framework beyond a purely observational tool and toward a structure that can support investment decision making. The modular design of the system, which separates news ingestion, sentiment computation, summarization, portfolio aggregation, and front-end rendering, provides a flexible foundation for further development. It also

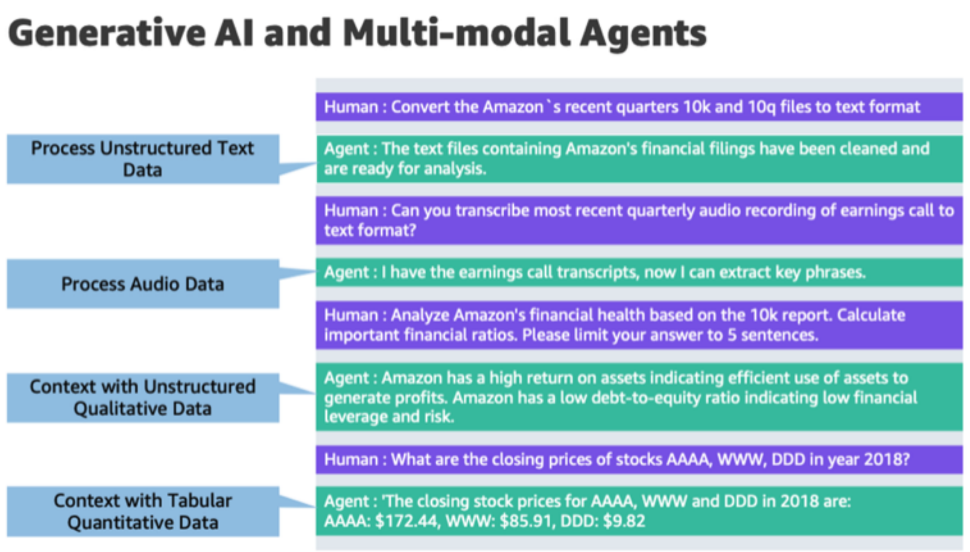


Figure 17: This figure, adapted from AWS (2023), illustrates how generative AI agents can process heterogeneous financial data—unstructured text such as 10-K and 10-Q filings, audio recordings from earnings calls, qualitative disclosures, and structured numerical data—to support complex analytical tasks. Each agent specializes in a distinct modality, enabling workflows such as document conversion, audio transcription, ratio extraction, qualitative interpretation, and quantitative reasoning

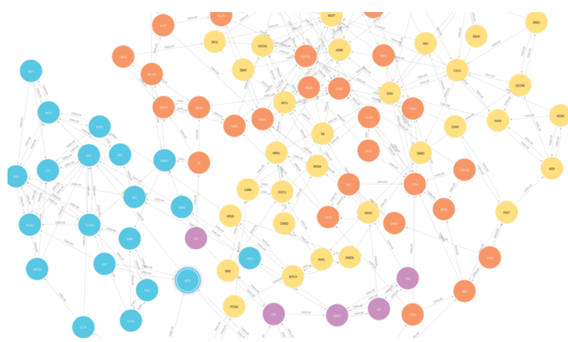

Figure 18: This figure, adapted from Neo4j (2023)'s graph analytics example, illustrates how financial entities—such as companies, themes, and co-mentioned events—can be represented as nodes in a heterogeneous graph. Edges encode relationships derived from co-occurrence patterns, shared risk exposures, or topical similarity. Such graph structures enable downstream tasks including identification of central nodes of influence, detection of sentiment spillovers across sectors or supply chains, and early recognition of emerging market clusters. Integrating graph neural networks or graph-ranking algorithms into the platform would significantly expand its analytical depth beyond linear sentiment aggregation, supporting richer inference over complex financial news networks.

illustrates how interpretable large language model outputs can complement quantitative sentiment scores and enhance both intuition and auditability.

Looking ahead, the system provides a base for more ambitious developments at the intersection of natural language processing, large language models, and investment analytics. Integrating additional data sources such as earnings call transcripts, regulatory filings, and structured fundamentals would enrich the understanding of firm-level narratives. The introduction of retrieval-augmented generation would allow the summary agent to anchor its explanations in verified disclosures rather than relying entirely on short-horizon news. Extending the existing portfolio-level module to incorporate sentiment spillover effects, cross-sectional interactions, and scenario evaluation would further align the platform with professional risk-management practice. As improvements accumulate in multi-modal reasoning, multi-agent coordination, and diversified data ecosystems, the system can evolve into a comprehensive artificial intelligence research assistant capable of supporting both fundamental research and sentiment-driven investment strategies.

## References


Baker, M., and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61(4), 1645–1680.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62(3), 1139–1168.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104(2), 288–302.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson (2013). Which news moves stock prices? A textual analysis. *Review of Financial Studies* 26(1), 21–81.

Fang, L., and J. Peress (2009). Media coverage and the cross-section of stock returns. *Journal of Finance* 64(5), 2023–2052.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *Journal of Finance* 66(5), 1461–1499.

Bollen, J., H. Mao, and X. Zeng (2011). Twitter mood predicts the stock market. *Journal of Computational Science* 2(1), 1–8.

Ranco, G., D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič (2015). The effects of Twitter sentiment on stock price returns. *PLOS ONE* 10(9), e0138441.

Loughran, T., and B. McDonald (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66(1), 35–65.

Malo, P., A. Sinha, P. Korhonen, J. Wallenius, and P. Takala (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology* 65(4), 782–796.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*.

Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint* arXiv:1908.10063.

Yang, Y., M. C. S. Uy, and A. H. Huang (2020). FinBERT: A pretrained language model for financial communications. *arXiv preprint* arXiv:2006.08097.

Lopez-Lira, A., and Y. Tang (2023). Can ChatGPT forecast stock price movements? Return predictability and large language models. *arXiv preprint* arXiv:2304.07619.

Xie, Q., W. Han, Y. Lai, M. Peng, and J. Huang (2023). The Wall Street neophyte: A zero-shot analysis of ChatGPT over multimodal stock movement prediction challenges. *arXiv preprint* arXiv:2304.05351.

AWS (2023). Generative AI and multi-modal agents in AWS: The key to unlocking new value in financial markets. *AWS Machine Learning Blog*. Available at: `https://aws.amazon.com/blogs/machine-learning/generative-ai-and-multi-modal-agents-in-aws-the-key-to-unlocking-new-value-in-financial-mar`

Kharshit (2024). RAG for financial reports using large language models. *Personal Blog*. Available at: https://kharshit.github.io/blog/2024/04/26/rag-financial-reports-llm.w

Neo4j (2023). Diversify your stock portfolio with graph analytics. *Neo4j Developer Blog*. Available at: https://neo4j.com/blog/developer/diversify-your-stock-portfolio-with-graph-analytics/.