King's Capital: Sentiment analysis on equities using Large Language Models

Jay Shah
Alexandros Taliotis
Quantitative Trading Division

Abstract

This paper explores the application of FinBERT, a financial adaptation of BERT, for sentiment analysis in financial markets. Built on the Transformer architecture, FinBERT is fine-tuned on financial texts to capture domain-specific language and sentiment nuances. At King's Capital, FinBERT was implemented to analyze equity news articles, providing sentiment scores to support trading strategies. Using data from CNBC and Yahoo Finance, headlines were preprocessed and analyzed, revealing predominantly positive sentiment for Apple-related news, with a significant neutral component. This study highlights the effectiveness of FinBERT in extracting actionable insights for financial decision-making.

Backround and Context

The financial world thrives on information. Market movements often hinge on how investors perceive the latest company developments, economic indicators, or geopolitical events. As the volume of digital news rapidly expands, sentiment analysis becomes a powerful tool for extracting actionable insights from textual data. By determining whether the prevailing tone of news coverage is positive, negative, or neutral, investors and analysts can gain a deeper understanding of market psychology, better manage risk, and potentially predict price movements.

In recent years, the use of large language models (LLMs) such as BERT (Bidirectional Encoder Representations from Transformers) has revolutionized Natural Language Processing (NLP). These models learn contextual relationships between words in a text by looking at both the left and right contexts simultaneously, leading to more accurate representations of language. In the finance domain, FinBERT, a variant of BERT specifically trained on financial texts, further refines language understanding for financial documents, news, and reports. This specialized model enables more precise sentiment analysis tailored to the nuances of financial language, jargon, and context.

Motivation

At the Quant Trading department of King's Capital, we have utilized FinBERT to obtain sentiment scores of equities based on articles mentioning specific stocks over a certain period. This process aims to support our paper trading activity and provide useful guidance to the equities and derivatives teams. Sentiment analysis will be used in conjunction with other strategies developed by the Quant Department and not as a standalone strategy as other factors besides sentiment can contribute to price fluctuations.

The key resources used were the papers "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Devlin et. al and "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models" by Araci. Both papers were studied and analyzed to understand the technology and be able to deploy it at scale. Fundamental to our understanding of attention and transformant which comprise the backbone of these models was the paper "Attention is all you need" by Vaswani et. al. All resources alongside a notebook containing our used code will be provided in the appendix.

Theoretical background: Explaining BERT and FinBERT

2.1.1 Introduction to the Transformer Architecture

BERT (Bidirectional Encoder Representations from Transformers) is built on the Transformer architecture, introduced by Vaswani et al. in the seminal paper "Attention is All You Need." Transformers rely on self-attention mechanisms to process entire sequences of tokens in parallel, rather than iterating over each token in a fixed order.

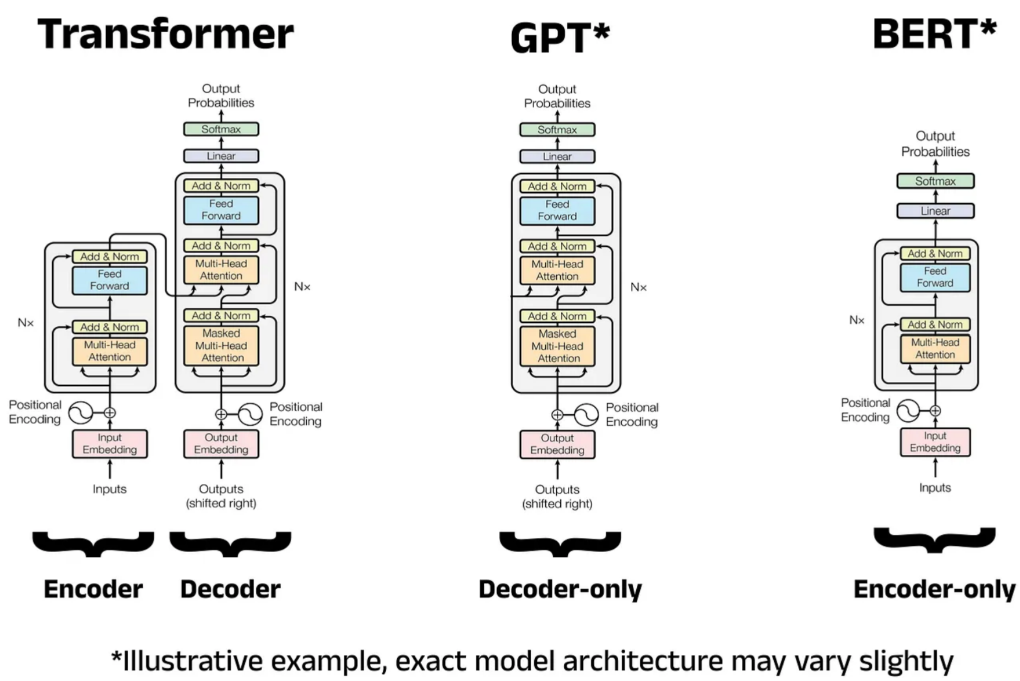
A Transformer is composed of two primary components: an encoder and a decoder. BERT specifically uses only the encoder part of the Transformer for its pre-training and fine-tuning tasks. Within each encoder layer, multiple self-attention heads allow the model to learn different ways of relating every word to every other word in a sentence (or even in larger contexts, depending on how the input is structured). This multi-head self-attention captures diverse contextual relationships, which is crucial for tasks such as sentiment analysis, natural language inference, question answering, and more. The model architecture is illustrated in the diagram below (Vaswani et al. 2018). Going into depth on the self-attention mechanism is outside the scope of this paper and we encourage you to read the papers outlined in the appendix for a greater understanding of these concepts.

2.1.2 Bidirectionality and Contextual Representations

One of the strengths of BERT is its bidirectional nature. Instead of reading text strictly left-to-right or right-to-left, BERT looks at the entire sequence of words at once, allowing it to build context-dependent embeddings. For example, the word *"bank"* may appear in different contexts: it can represent a "river bank" or a "financial institution." BERT discerns the intended meaning by attending to surrounding words such as *"river"* or *"loan."* This makes BERT highly effective for tasks requiring nuanced language understanding, including sentiment analysis where subtle cues can dramatically shift the sentiment from positive to negative or vice versa.

2.1.3 BERT Input Embeddings

BERT combines three embedding types for each input token to capture different aspects of context. First, token embeddings represent the meaning of individual tokens (e.g., subwords like "play" as shown in figure 2 taken by the paper of Devlin et. al 2019). Second, segment embeddings distinguish between sentence A and sentence B in tasks such as Next Sentence Prediction. Finally, **positional embeddings** encode the order of tokens, compensating for self-attention's lack of inherent sequence awareness. By summing these three components, BERT receives a holistic representation of each token's identity, segment role, and position in the sequence.



2.1.4 Pre-training Tasks (Masked Language Modeling and Next Sentence Prediction)

BERT undergoes a self-supervised pre-training phase on a large corpus of text (e.g., Wikipedia, Books Corpus). Two key tasks are typically used:

-Masked Language Modeling (MLM): A certain percentage of tokens are randomly masked, and the model is asked to predict those masked tokens. This helps the model learn contextual representations of words, based on both left and right context.

-Next Sentence Prediction (NSP): The model sees pairs of sentences and predicts whether the second sentence actually follows the first in the original text. This teaches BERT to understand relationships between sentences. Note that while the original BERT included a Next Sentence Prediction objective, some newer variants (e.g., RoBERTa) remove or replace NSP with other objectives like Sentence Order Prediction

After pre-training, BERT can be fine-tuned on various downstream tasks, such as sentiment analysis, question answering, or named entity recognition, by simply adding a small classification layer and adjusting the weights with typically fewer than a few epochs on labeled task-specific data.

2.2.1 Fin BERT: Why a Financial Adaptation?

While BERT excels in a wide range of general language tasks, the financial domain has a specialized vocabulary and context that may not be fully captured by generic corpora such as Wikipedia or everyday text. Terms like "bullish," "bearish," or "interest rate hike" have strong contextual and sentiment-oriented connotations specific to finance.

As a result, FinBERT is introduced to address these nuances. FinBERT is typically BERT's base architecture, but fine-tuned or pre-trained further on large financial text datasets. This specialized training allows it to understand domain-specific terminology and better capture the sentiment signals within financial narratives.

2.2.2 Training and Fine-Tuning FinBERT

Like its general-purpose counterpart, FinBERT also relies on the Transformer's self-attention layers and the same underlying bidirectional context modeling. However, it differs primarily in the data used for pre-training or subsequent fine-tuning.

During this process, the masked language modeling objective may be kept to retain BERT's ability to predict missing words, but with a focus on finance-related contexts. Additionally, tasks like sentiment analysis can leverage labeled financial datasets, where each sentence or headline is annotated as positive, negative, or neutral.

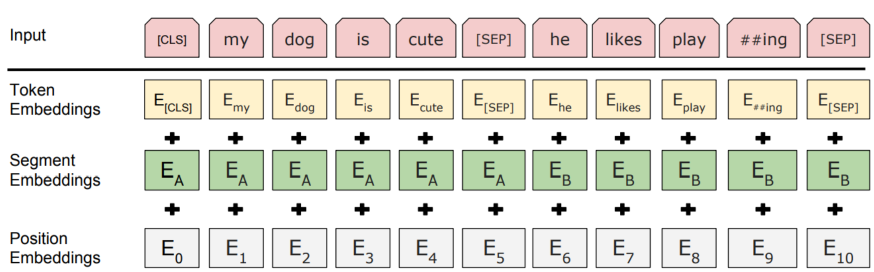


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

Implementation and results

Data collection:

Our workflow begins by gathering relevant news headlines for the target set of stocks, MLTX, NVDA, AAPL, GOOG, HUBS, INTC, MDC and LOT. We employ the requests library to connect to external sites and BeautifulSoup to parse the returned HTML. Specifically, two primary data sources are used:

- CNBC: We scrape headlines referencing these stocks from CNBC's quotes page.

- Yahoo Finance: We access Yahoo Finance's built-in news API through the yfinance library, extracting relevant article titles based on the short name or ticker symbol.

- Quarterly Reviews: When analyzing Apple specifically, we also fetch the company's newsroom post for quarterly reports, extracting the text with BeautifulSoup and filtering out irrelevant sections.

These steps ensure that a wide range of news sources and official statements are collected automatically, removing the need for manual curation.

Pre-processing:

Once the raw headlines are scraped, they are consolidated into a set structure in Python to remove duplicates. This helps ensure each headline is only analyzed once. We also define a minimum headline length (five characters) to discard obviously invalid entries such as empty strings or single-word placeholders. This pre-processing phase standardizes the input text, paving the way for a more robust sentiment analysis step.

Sentiment-analysis:

Through Hugging Face's transformers library, we import both the tokenizer and model. Each headline is tokenized and run through FinBERT, producing a probability distribution across three sentiment labels: positive, neutral, and negative. We then aggregate these probabilities with a simple weighting scheme (e.g., +1 for positive, 0 for neutral, 1 for negative) to arrive at a single "score" per headline. Finally, these scores are aggregated and normalized across all headlines to derive an overall sentiment percentage.

Key findings:

Overall, many of the stocks exhibit a substantial "neutral" share of sentiment, which can partially be explained by FinBERT's tendency to classify more ambiguous or informational headlines as neutral. In other words, articles that are factual rather than overtly optimistic or pessimistic may be "misrepresented" as neutral instead of slightly positive. This phenomenon is especially visible for MLTX and LOT, where neutral sentiment dominates more than half of the total score (about 60% for MLTX and nearly 70% for LOT). While that high neutral portion could mask some mild optimism, it nevertheless highlights that not all coverage is outright bullish or bearish.

Beyond the neutral skew, several distinct patterns emerge. For example, NVDA, GOOG, and MCD show relatively strong positive sentiment, with NVDA at 50%, GOOG at nearly 60%, and MCD close to 47%. This suggests these companies have more headlines framed in a hopeful or favorable light, likely due to discussions around product breakthroughs, market growth, or positive financial performance. Conversely, some tickers show a more balanced split between positive and neutral, such as AAPL (38% positive, 39% neutral) or HUBS (32% positive, 40% neutral), indicating headlines that, while not overtly negative, still express a fair amount of factual or wait-and-see coverage.

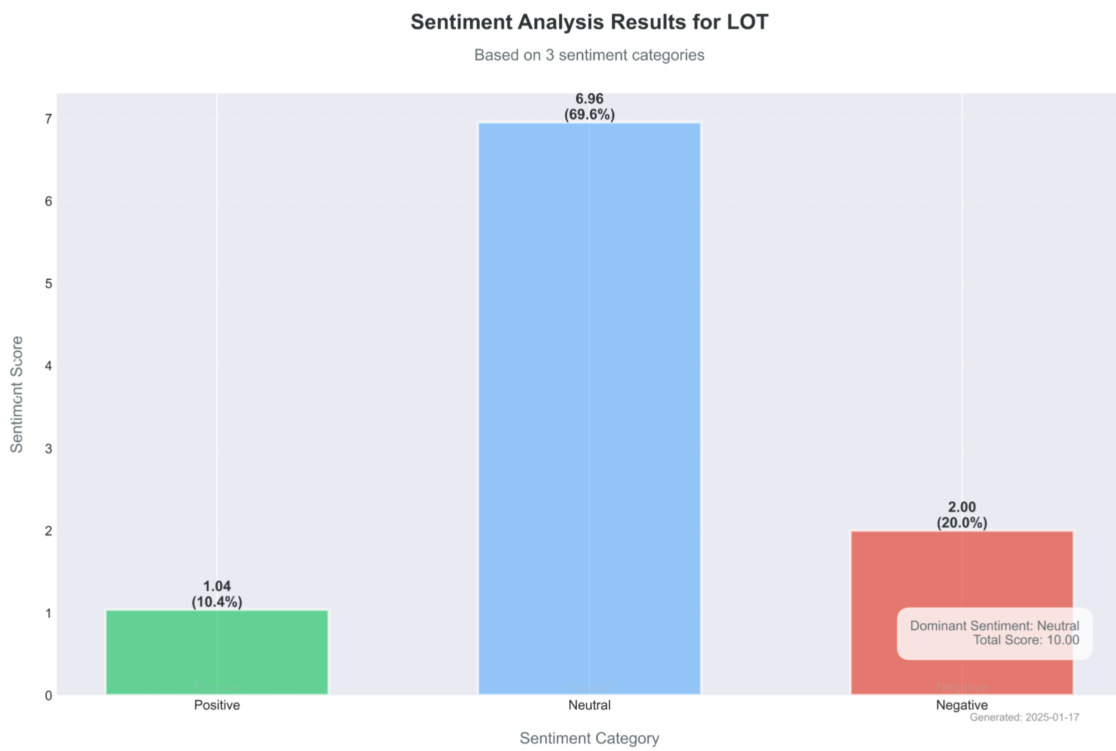
Negative sentiment generally remains lower across the board (e.g., 7% for MLTX, 9% for MCD), though a few symbols, like INTC (40.7% negative) and GOOG (38.7% negative), stand out with more critical or cautionary coverage. This could reflect skepticism regarding recent announcements or market headwinds. Still, even when negative sentiment is higher, the presence of a sizable neutral contingent hints that many headlines are nuanced or provide balanced reporting rather than clear condemnation.

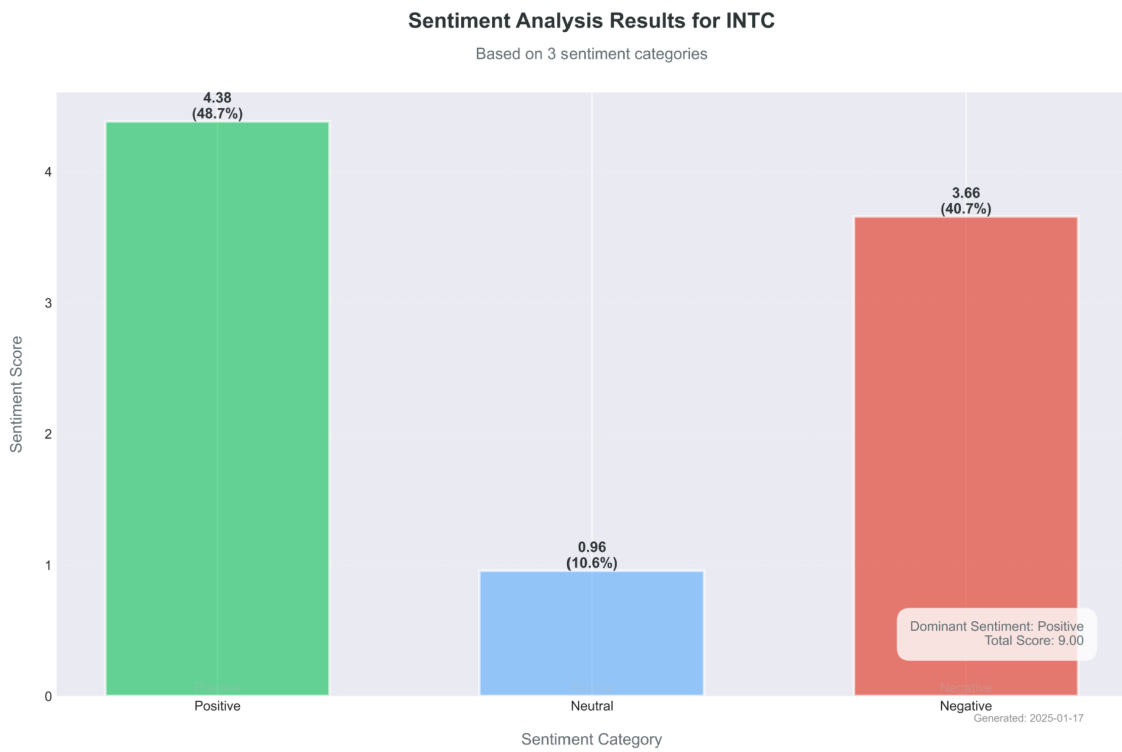
In practical terms, these mixed results remind us that FinBERT's classification can conflate mildly positive headlines with neutral ones, thus inflating the neutral category. Consequently, anyone using these scores for decision-making (e.g., gauging market outlook or investor sentiment) should account for the possibility that some "neutral" headlines may in fact be weakly positive. Overall, the data shows that most coverage tends toward either neutral or positive, with negative sentiment present but less dominant for most tickers, pointing to a generally constructive, though at times tempered, media landscape. The results of this paper are outlined in the Appendix section.

Appendix

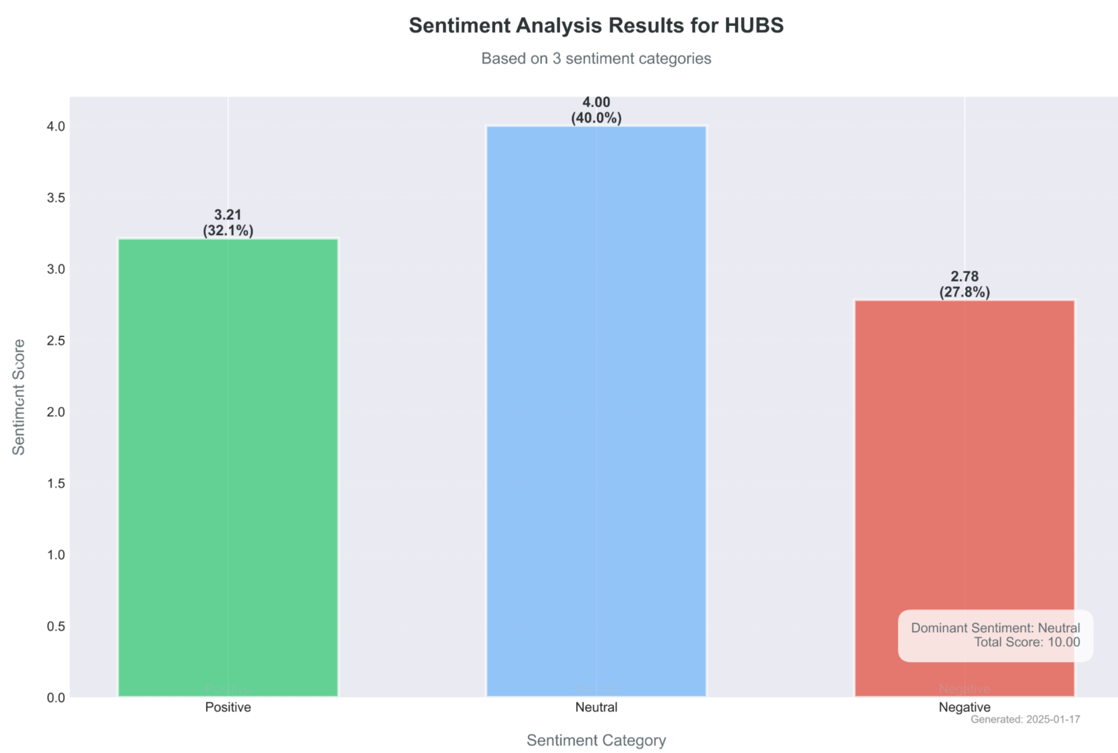
The GitHub repository containing our work can be accessed via the following link:
https://github.com/jays41/FinBERT_Paper

The following graphs illustrate the results of our experiment on our chosen set of stocks, discussed in detail in the **Key Findings** section. The articles are recent and the same time frame is used for all stocks in terms of article selection.

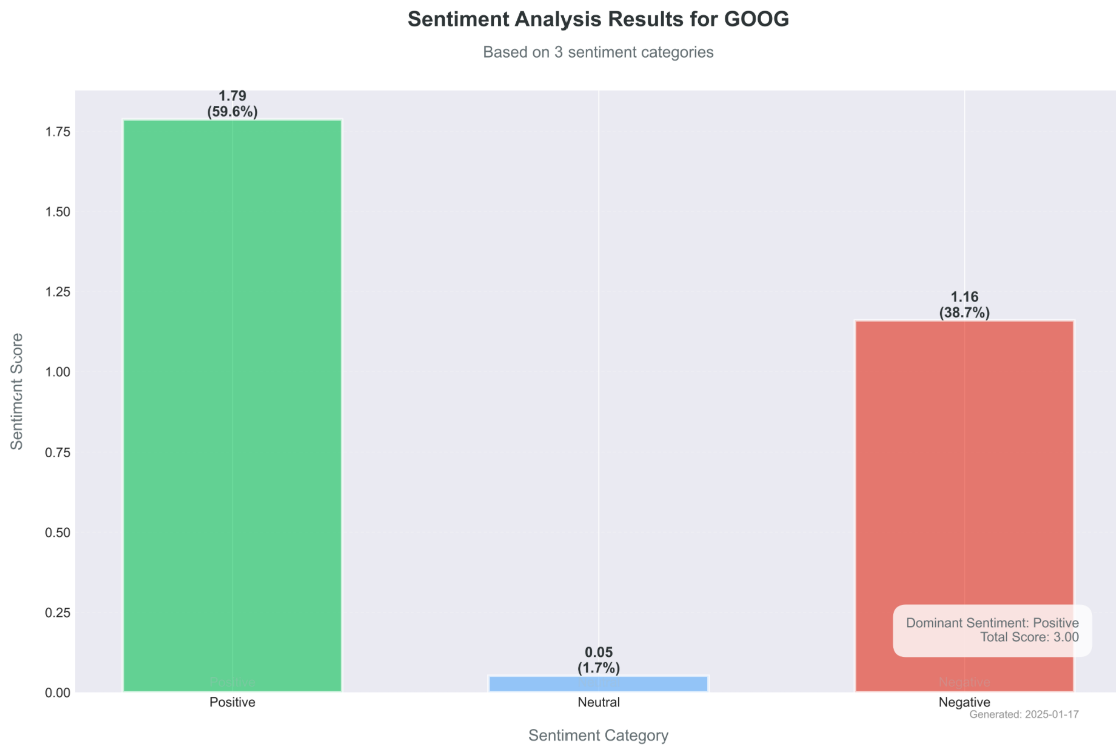




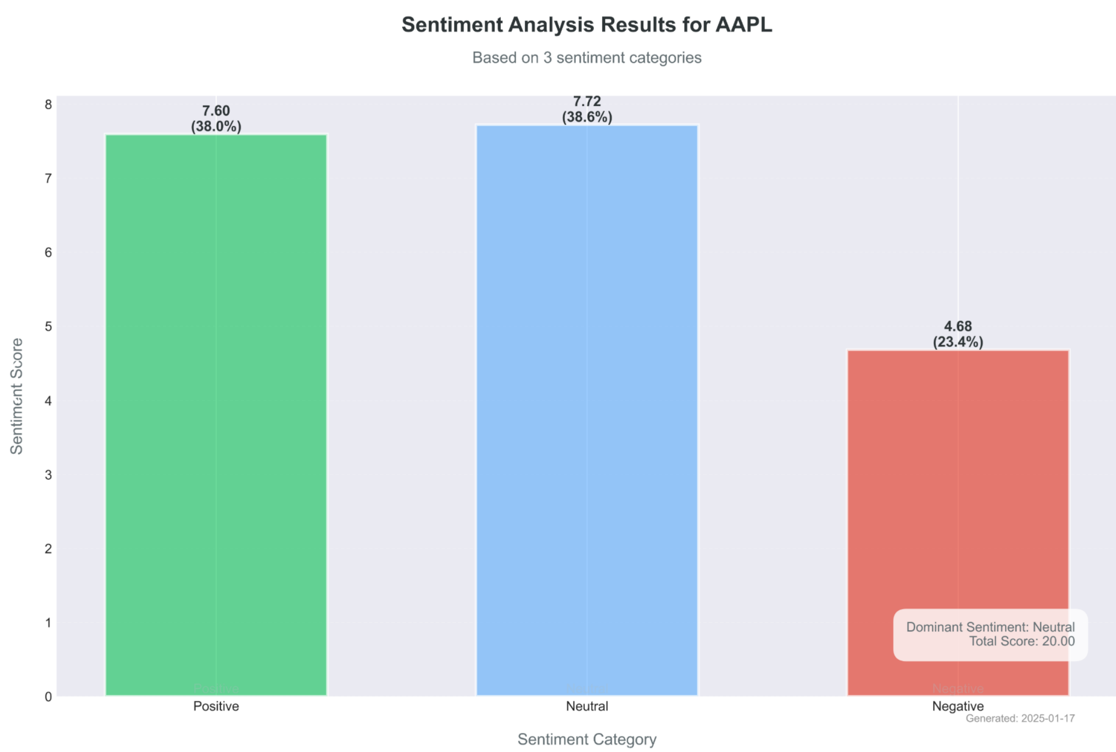
Sentiment Analysis Results for INTC
Based on 3 sentiment categories
4.38
(48.7%)
0.96
(10.6%)
3.66
(40.7%)
Sentiment Score
0
1
2
3
4
Positive
Neutral
Negative
Sentiment Category
Dominant Sentiment: Positive
Total Score: 9.00
Generated: 2025-01-17



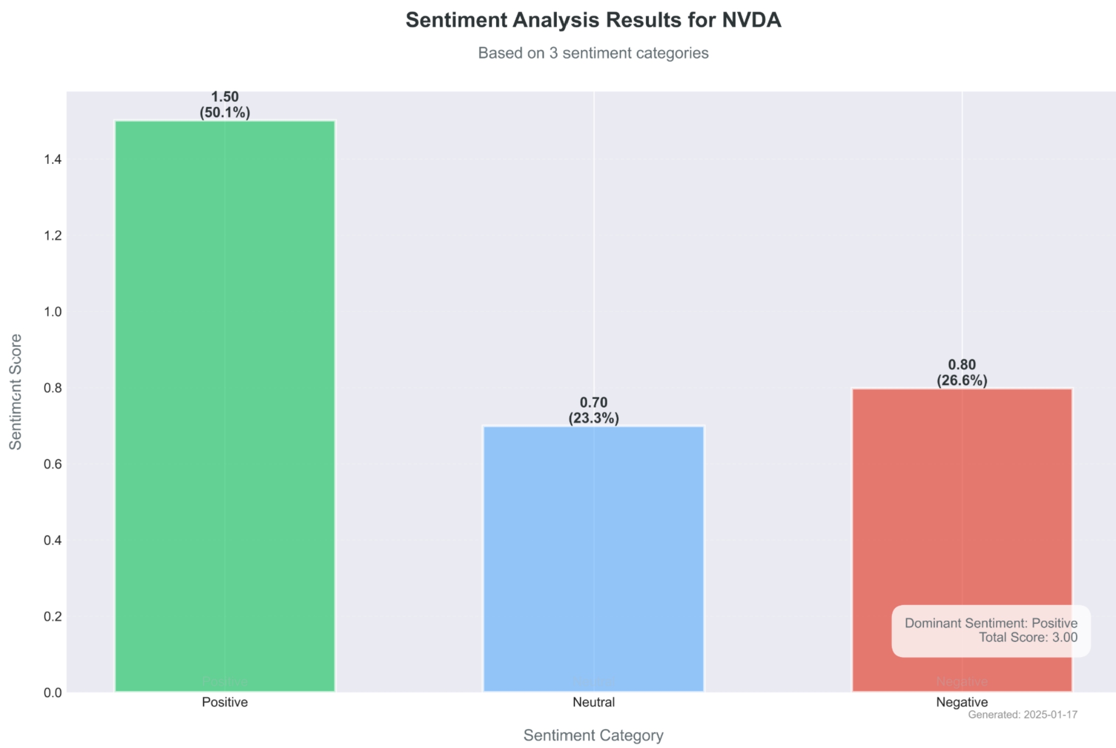
Sentiment Analysis Results for HUBS
Based on 3 sentiment categories
3.21
(32.1%)
4.00
(40.0%)
2.78
(27.8%)
Sentiment Score
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Positive
Neutral
Negative
Sentiment Category
Dominant Sentiment: Neutral
Total Score: 10.00
Generated: 2025-01-17

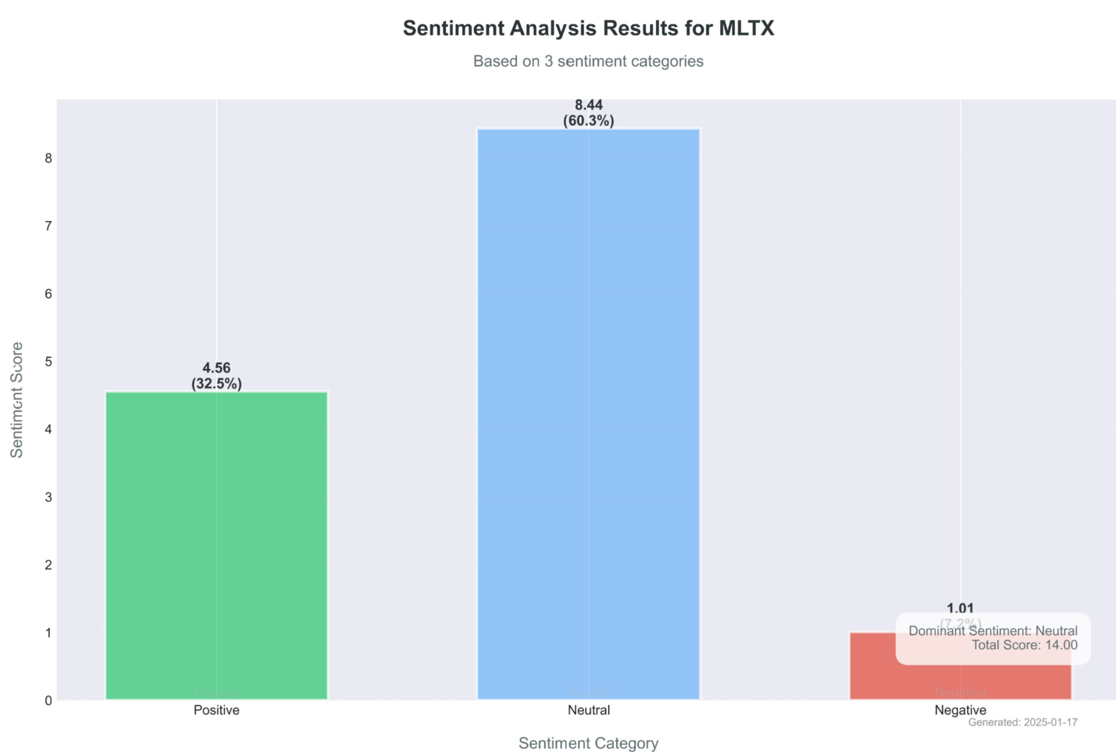


Sentiment Analysis Results for GOOG
Based on 3 sentiment categories
Sentiment Score
1.79
(59.6%)
0.05
(1.7%)
1.16
(38.7%)
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Dominant Sentiment: Positive
Total Score: 3.00
Positive
Neutral
Negative
Generated: 2025-01-17
Sentiment Category



Sentiment Analysis Results for AAPL
Based on 3 sentiment categories
Sentiment Score
7.60
(38.0%)
7.72
(38.6%)
4.68
(23.4%)
8
7
6
5
4
3
2
1
0
Dominant Sentiment: Neutral
Total Score: 20.00
Positive
Neutral
Negative
Generated: 2025-01-17
Sentiment Category


Sentiment Analysis Results for NVDA
Based on 3 sentiment categories
1.50 (50.1%)
0.70 (23.3%)
0.80 (26.6%)
Sentiment Score
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Positive
Neutral
Negative
Dominant Sentiment: Positive
Total Score: 3.00
Generated: 2025-01-17
Sentiment Category


Sentiment Analysis Results for MLTX
Based on 3 sentiment categories
4.56 (32.5%)
8.44 (60.3%)
1.01
Sentiment Score
0
1
2
3
4
5
6
7
8
Positive
Neutral
Negative
Dominant Sentiment: Neutral
Total Score: 14.00
Generated: 2025-01-17
Sentiment Category

Bibliography

Bibliography

Araci, D. T. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint* arXiv:1908.10063. Retrieved from https://arxiv.org/abs/1908.10063

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Retrieved from https://aclanthology.org/N19-1423/

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998–6008. Retrieved from https://papers.nips.cc/paper/7181-attention-is-all-you-need