BRAIN

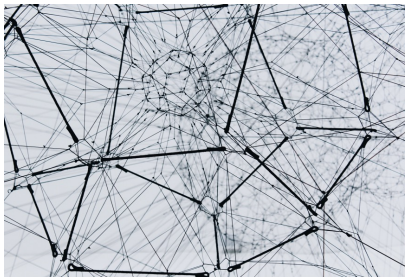
Brain News Topics Analysis with LLM

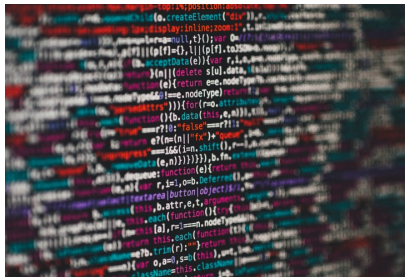
Use Cases

December 2024

Brain at a Glance

Brain is a research-driven company that specializes in developing customized solutions for investors and financial institutions using proprietary algorithms. Our innovative approach integrates **statistical methods**, **artificial intelligence**, and **natural language processing** to provide state-of-the-art solutions.

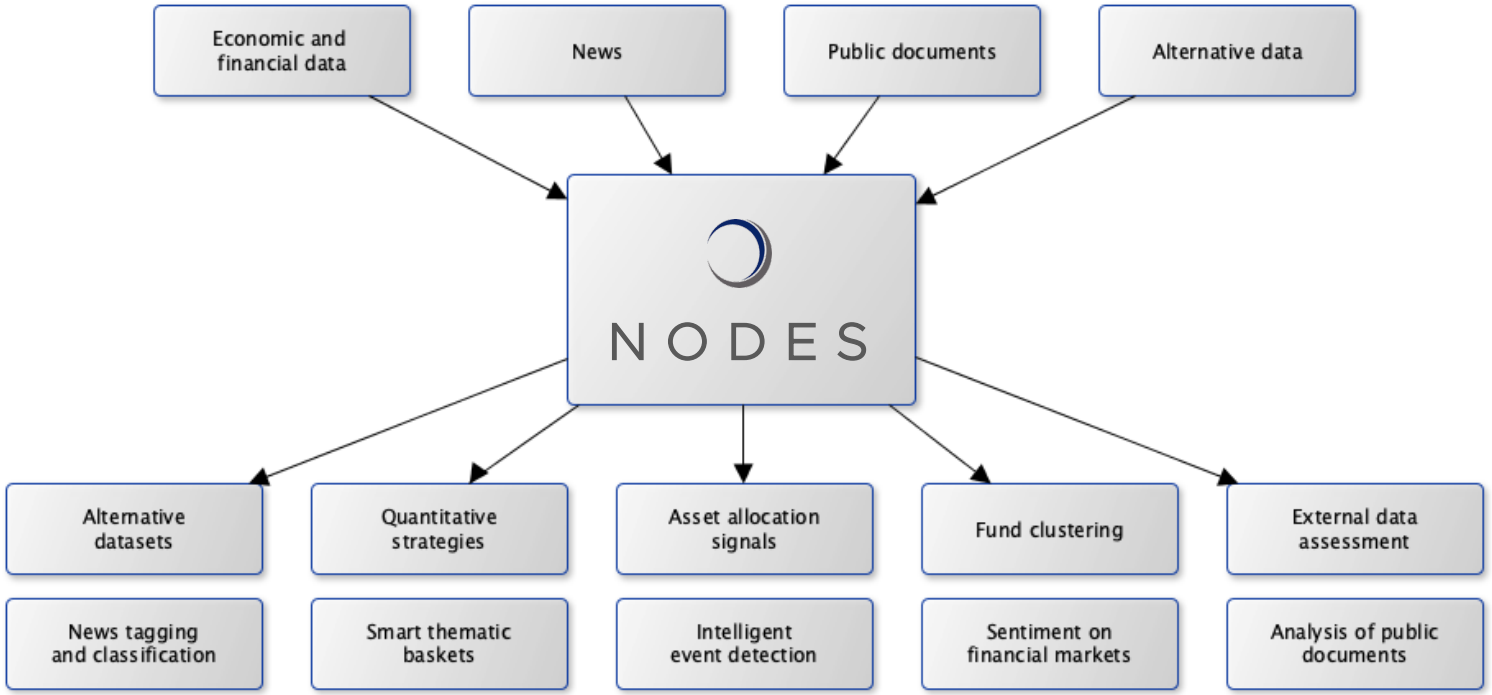




Our team is composed of experts with diverse academic backgrounds in **scientific research** and extensive **experience in financial markets**. This expertise allows us to support clients in the development, optimization, and validation of their **proprietary financial models**.

At Brain, our proprietary AI platform, **Nodes**, empowers clients to fully **leverage the potential of Artificial Intelligence and Large Language Models**. Nodes can process **financial and economic time series, alternative data, corporate and public documents** to individuate patterns and information related to financial markets.

NODES - Brain Artificial Intelligence Engine



With custom implementations of NODES, Brain is an **enabler** for the user to implement powerful solutions based on the state-of-the-art Artificial Intelligence and Large Language Models.

Clients and Commercial Partners

- **Quant hedge funds:** Brain is selling its proprietary datasets to large quant hedge funds with several billions of AUM, mostly US based. The typical client fund has large technological and datascience capabilities and is able to integrate many data sources.
- **Primary financial institutions:** Brain has recently developed A.I. driven solutions empowering the client to fully leverage the capabilities of A.I. and LLMs.
- **Data distribution platforms:** Brain has been selected by some well-known data distribution platforms to integrate and re-distribute its alternative data to quant and quantamental funds.

Bloomberg · Eagle Alpha · AWS Data Exchange · Datarade

- **Data visualization platforms:** data analytics platforms for the quantamental institutional investors are integrating Brain datasets to create signals and provide deeper insights to their clients. See for example **Point Focal.**
- **Fintech companies**: Brain datasets are provided to some fintech companies based in Italy and abroad operating in the asset management space.

Brain News Topics Analysis with LLM

Brain News Topics Analysis with LLM

- The Brain News Topics Analysis dataset exploits a **proprietary Large Language Model approach** to monitor specific topics and their sentiment within the financial news flow for stocks.
- For example, an investor may want to identify all news related to the topic "**innovation**" for a set of companies and track their sentiment with respect to each specific topic. Similarly, another investor can be interested in tracking all news related to the topic "**risks for the company**" and their sentiment.
- For **each stock** and **each topic** three metrics are provided using the news published within a given time interval:
    1. The **volume** of news relevant for the topic
    2. The **buzz**, which measures the variation in the amount of news that are published for each topic.
    3. A **sentiment score for the specific topic**, ranging from -1 to +1.
- All metrics are calculated based on the news published within the chosen time interval, e.g. the past 7 days for which the model has identified news to be relevant for a particular topic.
- The **calculation of historical data** from 2017 for Russell 1000 stocks **required significant computational power**, utilizing approximately 10 state-of-the-art GPUs running for three months on Brain's infrastructure.

Workflow

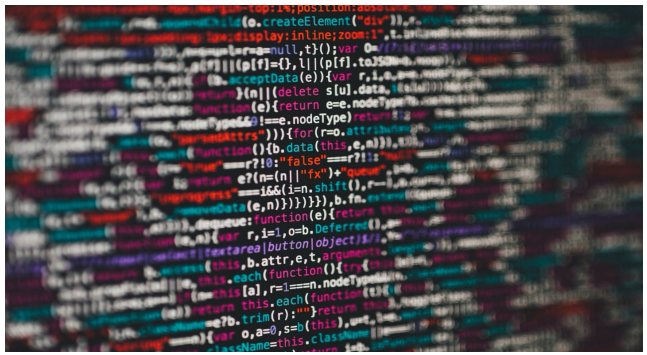
The Brain News Topics Analysis dataset exploits an internal and customized large language model to monitor specific topics and their sentiment within the financial news flow for stocks.

Data Collection and Tagging

Financial news is collected every few minutes from thousands of financial media sources via news aggregators that monitor both paid and free financial news sources and blogs for each stock. For example, in the first half of 2024, we tracked approximately 10,000 unique sources and collected an average of 20,000 daily news stories for Russell 1000 US stocks. News articles are tagged to each stock.

Topics Relevance and Sentiment

Brain Large Language Model evaluates the relevance of each news item to each of the 9 monitored topics. If a news item is considered relevant, the model assigns a sentiment score for that specific topic. This evaluation is based on the news headline. The LLM runs on Brain GPU cluster, ensuring full control over data privacy and model version management.



Data Aggregation

Daily, for each stock, **relevant news items for each topic are aggregated over various time periods** to calculate both a "buzz" score and a sentiment score for the topic. News repetition is taken into consideration during this phase. All metrics are calculated based on the news published within a given time interval, e.g. the past 7 days that the model identifies as relevant for each topic.

List of Monitored Topics

List of topics currently monitored by the Brain Large Language Model in the news flow:

1. Contracts, Licenses, and Partnership
2. Financial Results
3. Investor Asset Transactions and Positions
4. Governance and Management related Events
5. Innovation
6. Price variations
7. Rating and valuation estimates
8. Risks for the company
9. Legal

Example of news identified by the model as relevant to certain topics, along with their associated sentiments. The referenced stock is marked as "ASSET" in the headline.

Innovation

The race to dominate AI : Google and ASSET leading as computing costs surge.	POS
ASSET delays its EV plans, will keep making combustion engines.	NEG
ASSET is hoping its [...] computer headset, [...] but will we all wake around wearing augmented reality googles?	NEU

Rating and valuation estimates

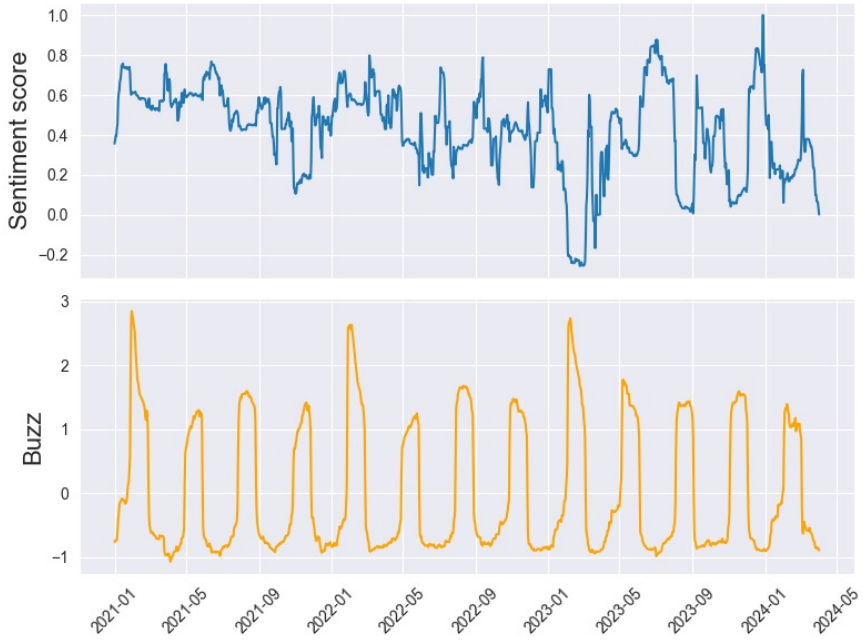
ASSET gets technical rating upgrade.	POS
ASSET rating lowered to market perform at Oppenheimer.	NEG
ASSET stock analysis : buy, sell, or hold in 2024?	NEU

Examples of Aggregated Metrics by Topic - 1/2

Aggregate sentiment and buzz for topic "innovation" for AAPL over the past 90-day period.

We observe that the sentiment on the topic of innovation is **biased towards the positive.**

Aggregate sentiment and buzz for topic "financial results" for AAPL over the past 90-day period.



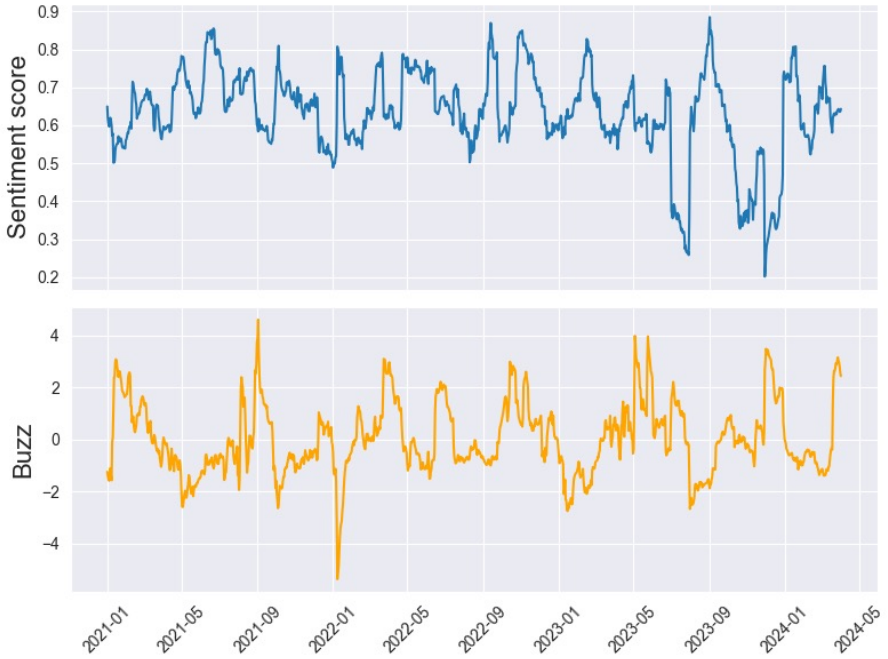
We observe that the buzz around the topic "financial results" shows quarterly spikes corresponding to earnings releases.

Examples of Aggregated Metrics by Topic - 2/2

Aggregate sentiment and buzz for topic "risks for the company" for AAPL over the past 30-day period.

We observe that the sentiment on the topic "risks for the company" innovation is **biased towards the negative.**

Aggregate sentiment and buzz for topic "contracts, licenses and partnerships" for AAPL over the past 90-day period.



We observe that the sentiment on the topic "contracts, licenses and partnerships" innovation is **biased towards the positive.**

Brain News Topics Analysis with LLM (BNTA_LLM)

Main Objective	The Brain News Topics Analysis dataset exploits an internal large language model to monitor specific topics and their sentiment within the financial news flow for stocks.
Covered universe	Russell 1000 stocks
History	8 years, starting from 2017
Update frequency	Daily updates (csv files delivered by 5 a.m. EST to be used for trading on same day)
Lag of updates	In the data delivered on date D, all news up to and including date D-1 are included in the analysis.
Main technology	Natural Language Processing and **proprietary Large Language Model** approach.
Why interesting to a systematic PM	Large language models have demonstrated high accuracy in extracting specific topics from news and can be utilized to derive structured data for systematic strategies from unstructured textual sources. The dataset metrics can be used stand alone to rank stocks for an investment strategy or combined in a more complex ML model to predict the stock ranking on various future time-horizons.
Dataset metrics	For each stock and each topic three metrics are provided using the news published within a given past time interval (previous day, past 7, 30,90 and 180 days): • The volume of news relevant for the topic • The buzz, which measures the variation in the amount of news that are published for each topic. • A sentiment score for the specific topic, ranging from -1 to +1.
Monitored topics	**9 financial topics** are currently monitored: Contracts, Licenses, and Partnerships; Financial Results; Investor Asset Transactions and Positions; Governance and Management related Events; Innovation; Price Variations; Rating and Valuation Estimates; Risks for the Company; Legal.
Dataset sharing	The dataset files can be shared via FTP or an AWS S3 bucket.

Dataset Structure

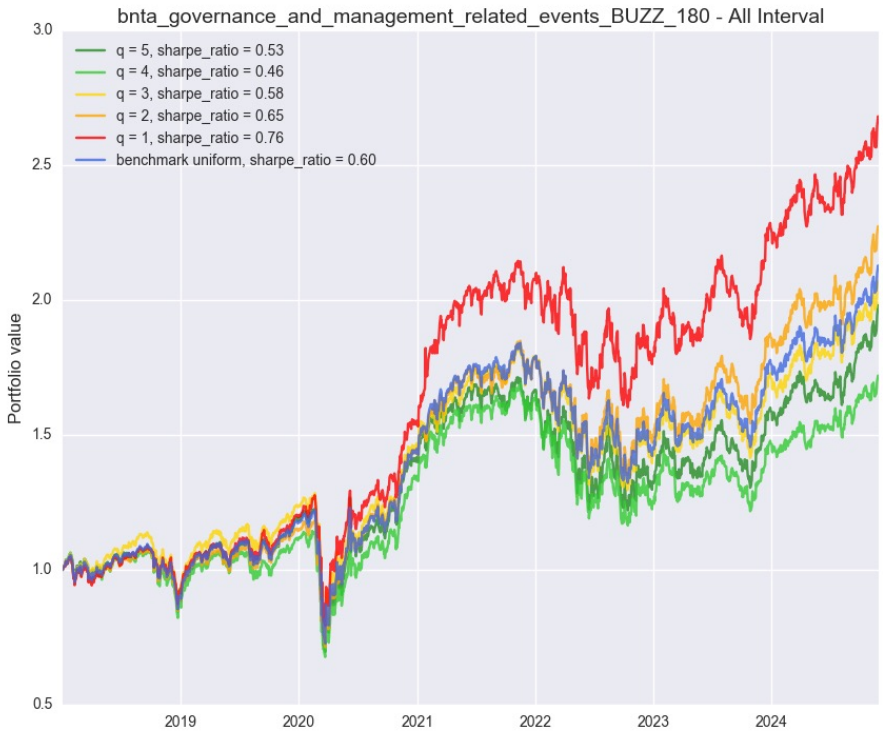
The dataset has the following structure:

- **Calculation date:** data labelled with date D are published within 8 a.m. UTC so it can be used for trading on the same day. All news up to and including date D-1 is included in the analysis.
- **Stock identifiers:** the primary unique identifier is the FIGI code but also the *ticker* is provided
- **Metric columns** are labelled as $TOPIC_NAME_$METRIC_$TIME_HORIZON, representing the aggregate $METRIC (sentiment, volume, buzz) for the topic $TOPIC_NAME over the past $TIME_HORIZON days. If no value is present, it indicates that no relevant financial news for this topic was identified during the aggregation period.

DATE	COMPOSITE_FIGI	TICKER	INNOVATION_SENT_90	INNOVATION_VOLUME_90	INNOVATION_BUZZ_90	RISKS_FOR_THE_COMPANY_SENT_90	-----
03/06/24	BBG000B9XRY4	AAPL	0,7673	2937	1,346	-0,5354	.....
03/06/24	BBG000BBS2Y0	AMGN	0,5047	50	0,2062	-0,1474	.....
03/06/24	BBG000BVPV84	AMZN	0,7306	54	0,1032	-0,1002	.....
03/06/24	BBG000BCQZS4	AXP	0,6395	56	0,9698	-0,1068	.....
03/06/24	BBG000BCSST7	BA	0,5004	439	2,641	-0,4977	.....
03/06/24	BBG000BF0K17	CAT	0,7772	40	-1,2146	-0,4963	.....
03/06/24	BBG000BN2DC2	CRM	0,8064	193	0,6052	-0,5236	.....
03/06/24	BBG000C3J3C9	CSCO	0,7373	269	0,3838	-0,261	.....
03/06/24	BBG000K4ND22	CVX	0,635	85	0,7624	-0,4634	.....
03/06/24	BBG000BH4R78	DIS	0,6086	63	-2,1005	-0,4938	.....
03/06/24	BBG00BN96922	DOW	0,6274	61	2,4017	-0,0114	.....
03/06/24	BBG000H556T9	HON	0,6281	26	-1,5307	0,0671	.....
03/06/24	BBG000BLNNH6	IBM	0,6561	434	-0,6793	-0,3422	.....
03/06/24	BBG000BWXBC2	WMT	0,71	163	-0,6691	-0,5158	.....
03/06/24	.....	.....	.....	.....	.....	.....	.....

Quintile Analysis of Buzz for Topic "Governance and Management Relates Events"

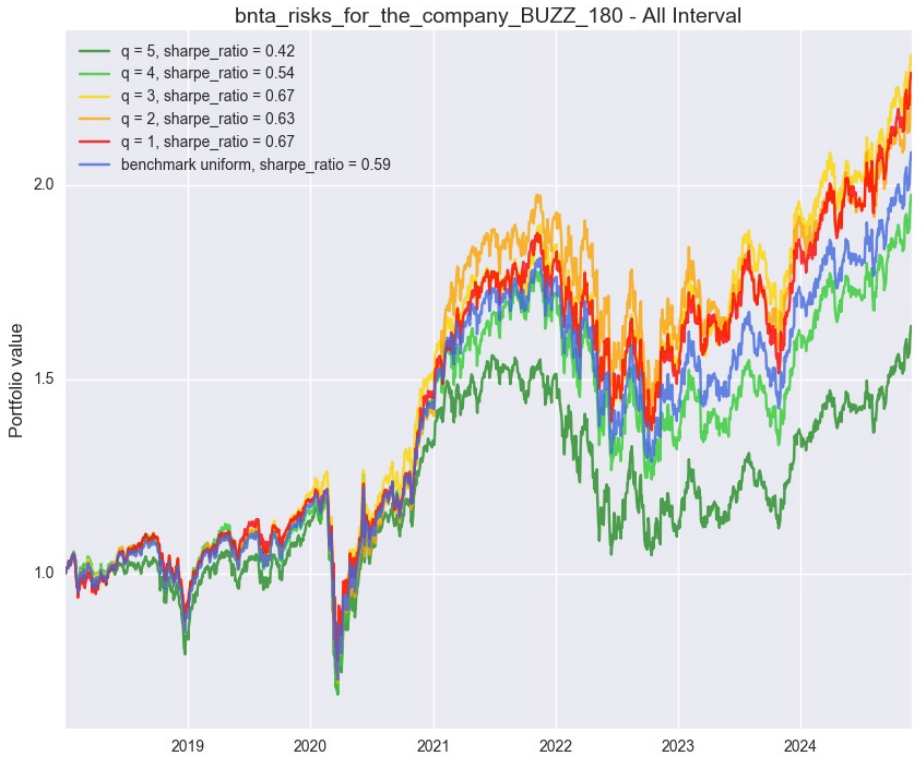
Summary / General Idea	Stocks with lower buzz (fewer spikes in news volume) related to "Governance and Management" events over the previous six months outperform stocks with higher buzz. One possible explanation is that higher buzz is often linked to negative events or scandals involving management, which can adversely impact stock performance.
Validation Approach	At each rebalancing event, stocks are divided into quintiles based on the buzz value for the topic.
Dataset Field	GOVERNANCE_...BUZZ_180 in file "metrics"
Universe	Stocks in the dynamic universe made of 1000 most liquid US stocks, approximately corresponding to the Russell 1000 index. The universe is updated each year using the most liquid stocks of previous year to avoid survival bias.
Interval	January 2018 – November 2024
Setup Details	Monthly rebalancing frequency, uniform weights, no commissions.
Results	The bottom quintile, representing stocks with lower buzz (red line, Sharpe ratio of 0.90) on the topic, outperforms the top quintile (dark green line, Sharpe ratio of 0.53), which corresponds to stocks with higher buzz. Additionally, the performance of the five quintiles approximately aligns in order with their ranking.



Past returns are not indicative of future performance.

Quintile Analysis of Buzz for Topic "Risks for the Company"

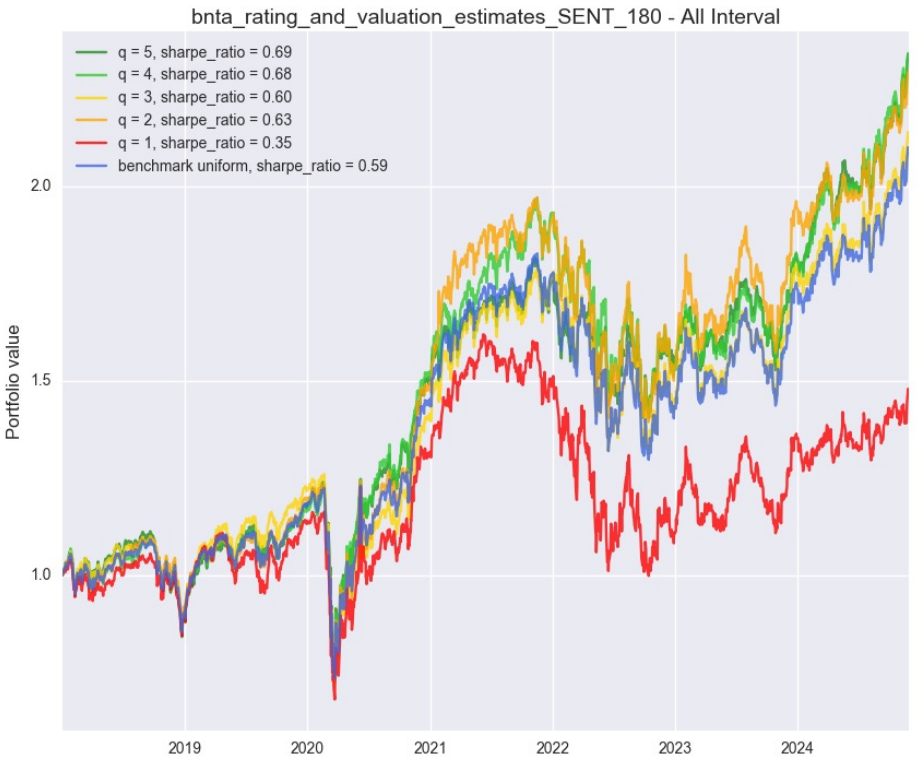
Summary / General Idea	Stocks with higher buzz (characterized by larger spikes in news volume) related to "risks for the company" events over the past six months tend to underperform those with lower buzz. One possible explanation is that higher buzz is often associated with newly emerging risks for the company.
Validation Approach	At each rebalancing event, stocks are divided into quintiles based on the buzz value for the topic.
Dataset Field	RISKS_FOR_THE_COMPANY_BUZZ_180 in file "metrics"
Universe	Stocks in the dynamic universe made of 1000 most liquid US stocks, approximately corresponding to the Russell 1000 index. The universe is updated every year using the most liquid stocks from the previous year to avoid survival bias.
Interval	January 2018 – November 2024
Setup Details	Monthly rebalancing frequency, uniform weights, no commissions.
Results	The top quintile, representing stocks with higher buzz (green line, Sharpe ratio of 0.42) on the topic of "risks for the company," significantly underperforms compared to the other quintiles. For instance, the bottom quintile (red line, Sharpe ratio of 0.67) represents stocks with lower buzz. Moreover, the performance of the five quintiles generally aligns with their rankings.



Past returns are not indicative of future performance.

Quintile Analysis of Sentiment for Topic “Rating and Valuation Estimates”

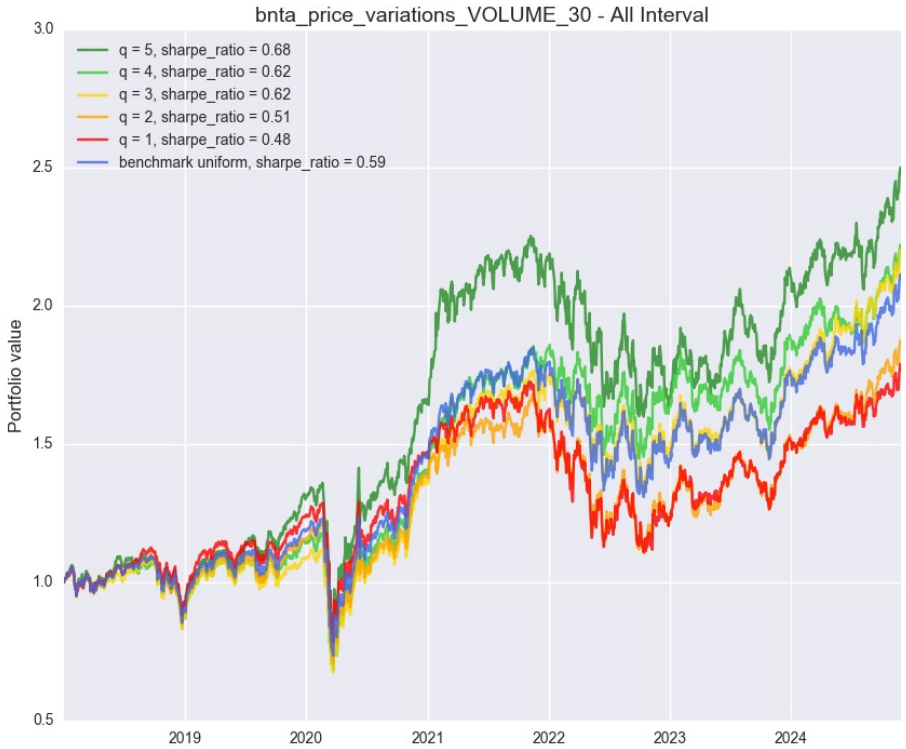
Summary / General Idea	Stocks with the lowest sentiment on the topic of “rating and valuation estimates” over the previous month underperform stock with higher sentiment. It seems that news with low sentiment regarding the company's rating and valuation leads to weaker performance of the company.
Validation Approach	At each rebalancing event, stocks are divided into quintiles based on the sentiment for the topic.
Dataset Field	RATING_AND_VALUATION_...SENT_30 in file ”metrics”
Universe	Stocks in the dynamic universe made of 1000 most liquid US stocks, approximately corresponding to the Russell 1000 index. The universe is updated each year using the most liquid stocks of previous year to avoid survival bias.
Interval	January 2018 – November 2024
Setup Details	Monthly rebalancing frequency, uniform weights, no commissions.
Results	The bottom quintile, representing stocks with the lowest sentiment (red line, Sharpe ratio of 0.35) on the topic of “rating and valuation estimates” underperforms compared to the other quintiles. For example, the top quintile (dark green line, Sharpe ratio of 0.69) represents stocks with the highest sentiment on this topic.



Past returns are not indicative of future performance.

Quintile Analysis of News Volume for Topic “Price Variations”

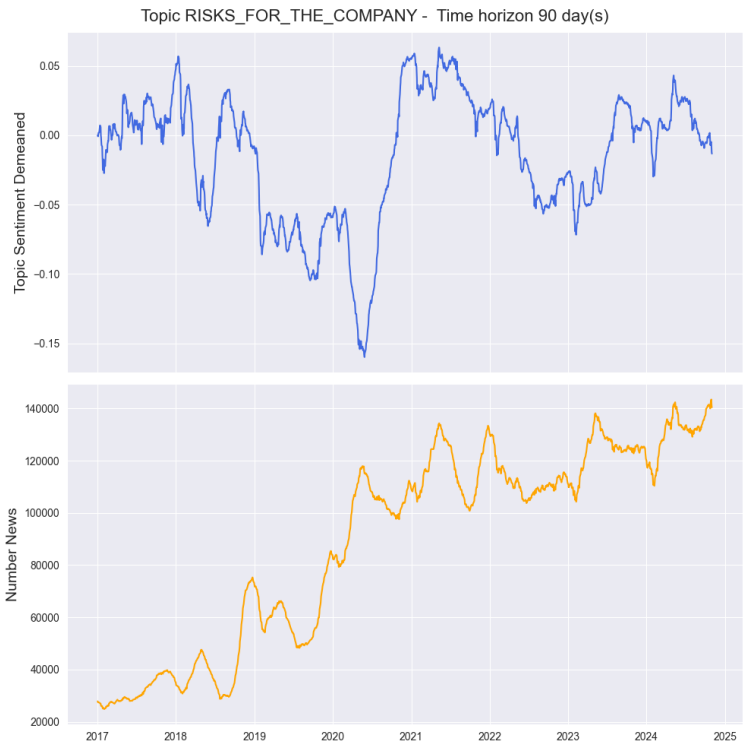
Summary / General Idea	Stocks with higher news volume related to “price variations” over the past month outperform those with lower news volume. A possible explanation is that higher news volume on “price variations” often indicates greater positive attention to the company, which may generally correspond to a favourable perception of the company by investors in the following period.
Validation Approach	At each rebalancing event, stocks are divided into quintiles based on the news volume for the topic.
Dataset Field	PRICE_VARIATIONS_VOLUME_30 in file ”metrics”
Universe	Stocks in the dynamic universe made of 1000 most liquid US stocks, approximately corresponding to the Russell 1000 index. The universe is updated each year using the most liquid stocks of previous year to avoid survival bias.
Interval	January 2018 – November 2024
Setup Details	Monthly rebalancing frequency, uniform weights, no commissions.
Results	The top quintile, representing stocks with the highest news volume (green line, Sharpe ratio of 0.68) on the topic of 'price variations,' outperforms the bottom quintile (red line, Sharpe ratio of 0.48), which represents stocks with the lowest news volume. Additionally, the performance of the five quintiles generally aligns with their rankings.

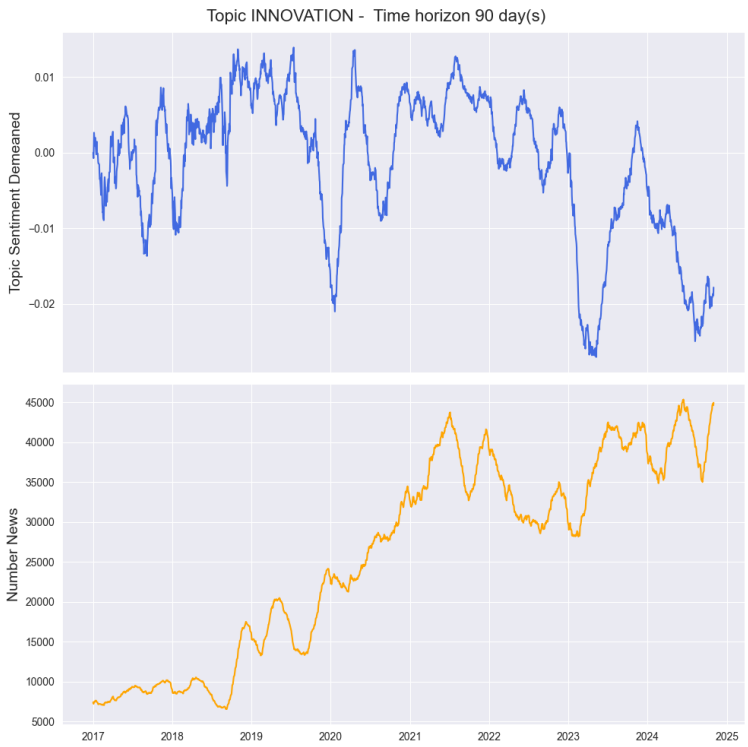


Past returns are not indicative of future performance.

Example of Aggregation at the Topic Level

The news analysis can be **aggregated at the topic level** to provide a general indication of sentiment and news volume for each topic. In the following plots, for two topics, we present the topic sentiment in the upper plot (mean subtracted from the beginning of the series) and the news volume within the 90 days time horizon in the lower plot. For the topic “risks for the company”, we observe, for instance, a decrease in sentiment during the first wave of COVID-19 at the beginning of 2020.





Possible Future Steps

- Inclusion of additional topics.
- Global coverage tailored to specific regions, aligned with client requests—for example, European stocks in the Eurostoxx 600 or other defined universes.
- Customization options to be discussed, considering the high computational power required.

All Brain Datasets

Currently Available Datasets

Single Ticker Datasets

Dataset	Coverage	Main Technology	Time Horizons	Update Freq.	History
Brain Sentiment Indicator	10.000 global stocks Sectors Forex, commodities, crypto	Natural Language Processing and Customized LLM	1, 7 and 30 past days	Daily	8+ years
Brain News Topics Analysis with LLM	Largest US stocks	Customized LLM	1, 7, 30, 90, 180 past days	Daily	7+ years
Machine Learning Stock Ranking	Largest US and EU stocks	Machine Learning	2, 3, 5, 10, 21 forward days	Daily	14+ years
Brain Language Metrics on Company Filings	6000+ US stocks	Natural Language Processing	Quarterly	Daily	14+ years
Brain Language Metrics on Earnings Calls Transcripts	4500+ US stocks	Natural Language Processing	Quarterly	Daily	12+ years
Brain Wikipedia Page Views	Largest 1000 US Stocks	Statistical Metrics	1, 7 and 30 past days	Daily	8+ Years
Machine Learning Stock Ranking on BLMCF features	Most liquid 3000+ US stocks	ML on features from NLP on filings	21 and 63 forward days	Daily	12+ years

Asset Allocation Signals

Dataset	Main Technology	Time Horizons	Update Frequency	History
Brain Market Sentiment	Natural Language Processing	1, 7 and 30 days	Daily and Intraday	6+ years
Dynamic Volatility Signal	Statistical model based on financial stress with focus on VIX	Few switches per year	Daily	16+ years

Disclaimer

The content of this presentation is not to be intended as investment advice. The material is provided for informational purposes only and does not constitute an offer to sell, a solicitation to buy, or a recommendation or endorsement for any security or strategy, nor does it constitute an offer to provide investment advisory or other services by Brain. Brain makes no guarantees regarding the accuracy and completeness of the information expressed in this document. Past returns are not indicative of future performance.

Contacts

BRAIN

- Web site: braincompany.co
- Email: contact@braincompany.co
- Linkedin page: Brain page