Brain Language Metrics on Global Earnings Calls Transcripts.

Product Summary

The usage of unstructured textual content, such as news, company filings, and earnings call transcripts (ECTs) is quickly expanding across both quantitative and discretionary financial strategies as reflected in the growing number of academic papers and commercial products in this domain.

The Brain Language Metrics on Earnings Calls Transcripts for Global Stocks (BLMECT_GLOB) dataset is one of the several datasets built by Brain with the objective of monitoring a large set of quantitative language metrics on documental information available for single stocks. This dataset is targeted at the quarterly ECTs of approximately 7000 global stocks. The ECTs are sourced from **Aiera** (https://aiera.com), a global platform that monitors events and communications for a large universe of listed companies.

The dataset is composed of two parts. Part one includes several **language metrics** for the most recent ECTs of each stock, namely:

1. Financial sentiment calculated using **Brain proprietary Large Language Model** approach
2. Percentage of words belonging to financial domain classified by language types:
    - *"Constraining"* language
    - *"Litigious"* language
    - *"Uncertainty"* language
    - "*Complexity*" language
3. Readability score
4. Lexical metrics such as lexical density and richness of text
5. Text statistics such as the transcript length

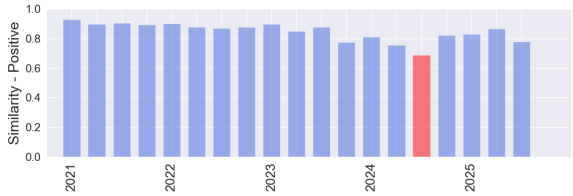
Part two includes the **differences** between the most recent earnings call transcript and the previous one:

1. Difference of the various language metrics (e.g. delta sentiment, delta readability score, delta percentage of a specific language type etc.)
2. Similarity metrics between documents, also with respect to a specific language type (for example similarity with respect to *"litigious"* language or "*uncertainty"* language)

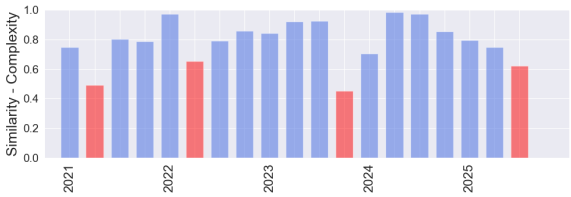
Examples of Metrics

One of the dataset metrics is the similarity between the most recent earnings call transcript and the previous one.

Similarity with focus on positive financial domain language for MSFT

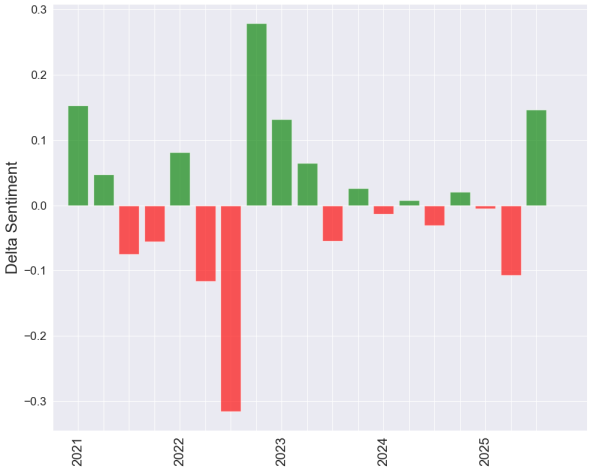


Similarity with focus on "complexity" financial domain language for NVDA



Another metric included in the dataset is the difference in sentiment scores between subsequent calls, capturing changes in the financial sentiment of the earnings calls.

Difference of Sentiment Score in NVDA Earnings Calls Transcripts.



Dataset Frequency and Delivery

The dataset is updated daily as new earnings call transcripts are published. For each stock, data is updated quarterly with new earnings calls. Historical data is available starting from 2021. The data is delivered daily as CSV files published in an S3 bucket.

Contacts

BRAIN is a research focused company that develops proprietary signals based on alternative data and algorithms for investment strategies on financial markets.

- EMAIL: contact@braincompany.co
- WEB: https://braincompany.co

Last update: 01/09/2025