DeltaSentiment: Sentiment analysis of 10-K reports using machine learning

Pataki Dávid, Pintér József

2024.06.11.

github.com/pinterj98/DeltaSentiment

MŰEGYETEM 1782

Input: US corporate 10-K reports (publicly available)

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 10-K

(Mark One)

☒ **ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**
For the fiscal year ended December 31, 2022
OR
☐ **TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**
For the transition period from _____ to _____
Commission File Number: 001-34756

Tesla, Inc.

(Exact name of registrant as specified in its charter)

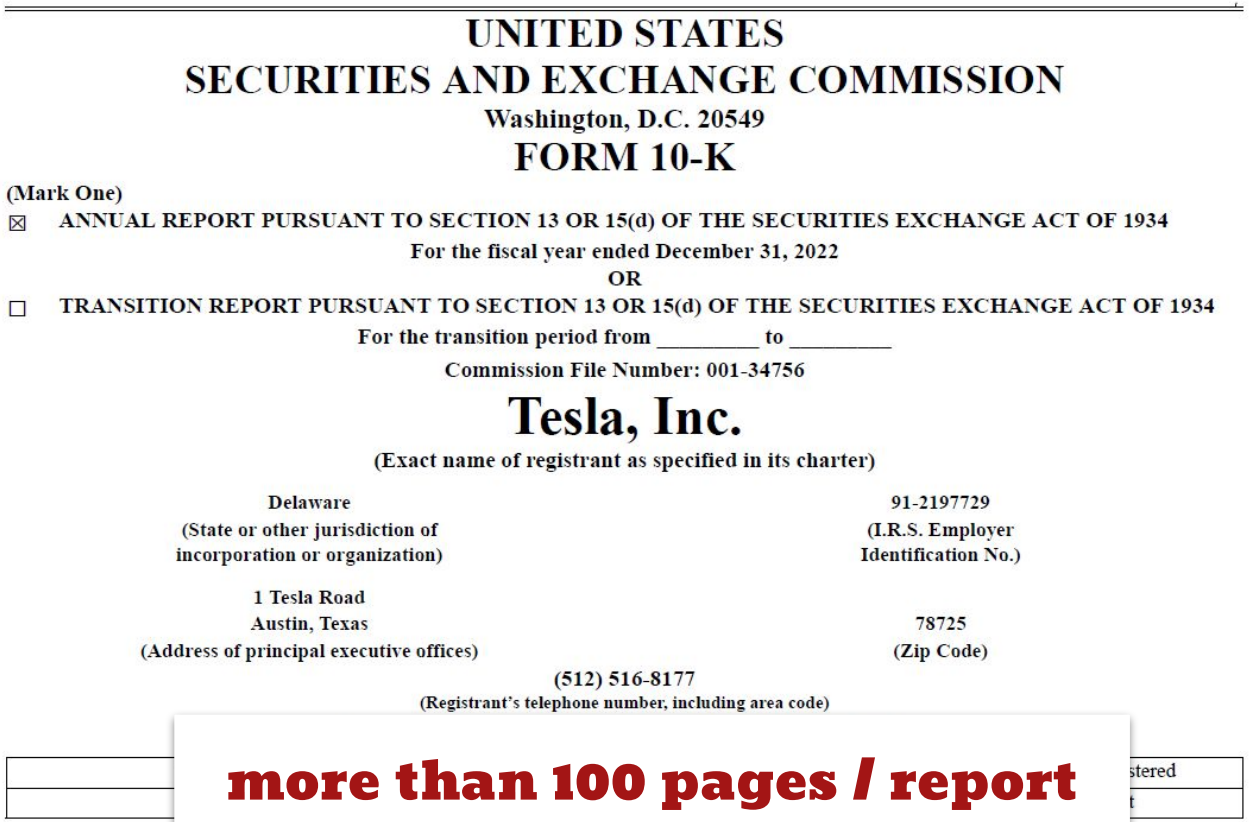
Delaware	**91-2197729**
(State or other jurisdiction of incorporation or organization)	**(I.R.S. Employer Identification No.)**
1 Tesla Road	
Austin, Texas	**78725**
(Address of principal executive offices)	**(Zip Code)**

(512) 516-8177
(Registrant's telephone number, including area code)

Securities registered pursuant to Section 12(b) of the Act:

Title of each class	Trading Symbol(s)	Name of each exchange on which registered
Common stock	TSLA	The Nasdaq Global Select Market

Input: US corporate 10-K reports (publicly available)



UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549
FORM 10-K

(Mark One)
☒ ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934
For the fiscal year ended December 31, 2022
OR
☐ TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934
For the transition period from _____ to _____
Commission File Number: 001-34756

Tesla, Inc.
(Exact name of registrant as specified in its charter)

Delaware	91-2197729
(State or other jurisdiction of incorporation or organization)	(I.R.S. Employer Identification No.)
1 Tesla Road Austin, Texas	78725
(Address of principal executive offices)	(Zip Code)

(512) 516-8177
(Registrant's telephone number, including area code)

more than 100 pages / report

Input: US corporate 10-K reports (publicly available)

Question: Can we obtain sentiment scores (😔/😶/😟), and “measure” how do they change from 2022 to 2023?

Input: US corporate 10-K reports (publicly available)

Question: Can we obtain sentiment scores (😌/😶/😟), and "measure" how do they change from 2022 to 2023?

more than 100 pages / report

Models we used: finBERT, finVADER, GPT

Brute force: copy + paste ⇨ **not feasible**

- finBERT has maximal context length of 512 tokens
- finVADER, LLM makes no sense for huge input

Input: US corporate 10-K reports (publicly available)

Question: Can we obtain sentiment scores (😌/😶/😟), and "measure" how do they change from 2022 to 2023?

more than 100 pages / report

Models we used: finBERT, finVADER, GPT

1. Convert pdf to text file
2. Select relevant sentences
3. Calculate sentiment score sentence by sentence

Preprocessing the documents

It is necessary to extract the relevant information from the documents.

1. Convert the pdfs into json text data.

Popular Python libraries
Can be fully customized
Not so good off the shelf performance

Java software
Python API
Ok off the shelf performance

Adobe Acrobat Pro software
Adobe API from Python
Good off the shelf performance

We use Adobe to convert pdf into html files.

Preprocessing the documents

It is necessary to extract the relevant information from the documents.

1. Convert the pdfs into json text data.

Convert pdfs to html files

Extract paragraphs from the documents

Split paragraphs into sentences

Raw PDF → List of sentences

Preprocessing the documents

It is necessary to extract the relevant information from the documents.

2. Extract relevant sentences
    a. Semantic Search
    b. Large Language Model

LLMs performed very differently

GPT - 4o > GPT - 3.5 - turbo ≈ Gemini Pro 1.0 > ~~Gemma 2b~~

GPT - 4o decides which sentences are important.

Preprocessing the documents

It is necessary to extract the relevant information from the documents.

2. Extract relevant sentences
    a. Semantic Search
    b. Large Language Model

LLMs performed very differently

GPT - 4o > GPT - 3.5 - turbo ≈ Gemini Pro 1.0 > ~~Gemma 2b~~

GPT - 4o decides which sentences are important.
Can we replace it with an independent smaller model?

Sentiment scores are calculated sentence by sentence

given combined_labeled_sentences.json for each report we do the following:

['sent_1','sent_2',...'sent_last'] →  → [score_1, score_2, ... score_last]

OR

$-1 \leq \text{score} \leq 1$

['sent_1','sent_2',...'sent_10'] → → [score_1, score_2,...score_10]

['sent_11','sent_12',...'sent_20'] → → [score_11, score_12,...score_20]

Sentiment scores are calculated sentence by sentence

given combined_labeled_sentences.json for each report we do the following:

['sent_1','sent_2',...'sent_last'] → 🤗 → ⟨ [score_1, score_2, ... score_last] ⟩

OR

$-1 \leq \text{score} \leq 1$

['sent_1','sent_2',...'sent_10'] →  → ⟨ [score_1, score_2,...score_10] ⟩

['sent_11','sent_12',...'sent_20'] →  → ⟨ [score_11, score_12,...score_20] ⟩

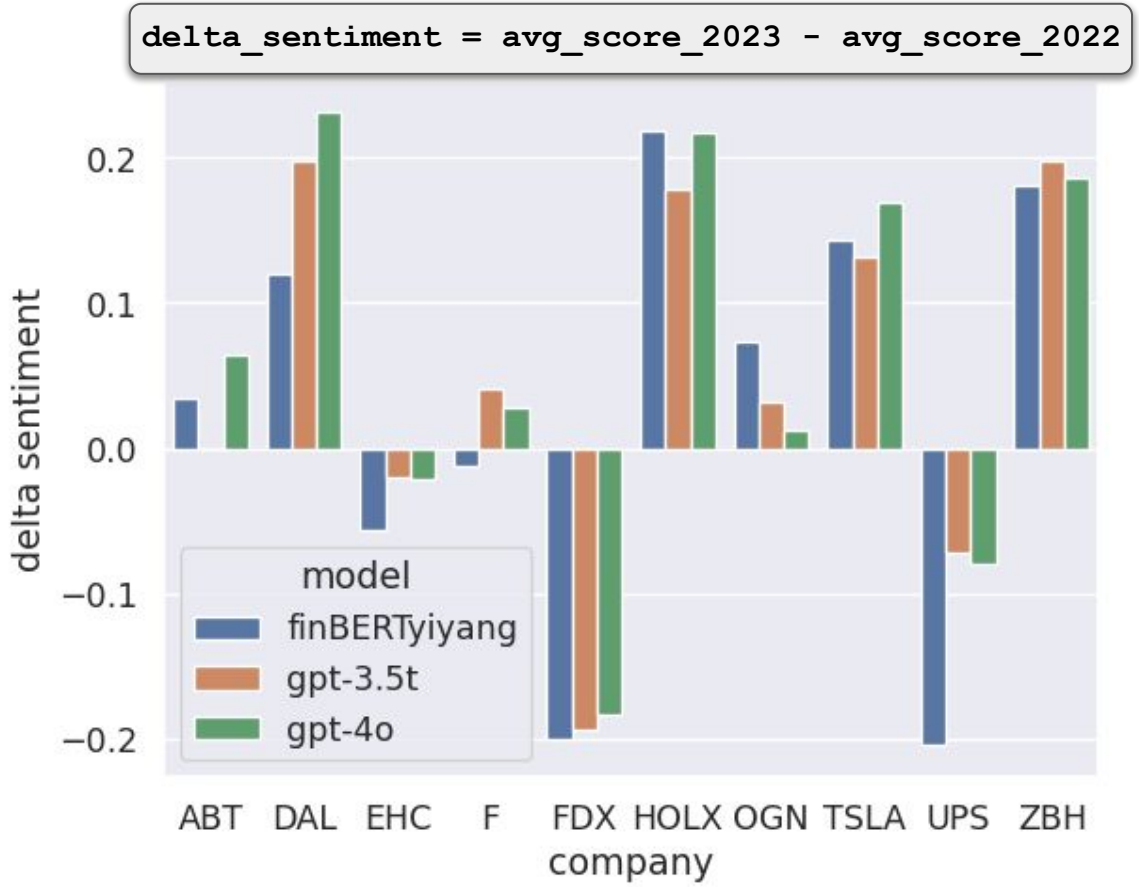
Results of finBERT and GPT show significant **correlation**

	finbertprosus	finbertyiyang	finvader	gpt-3.5t	gpt-4o
finbertprosus	1.000000	0.793237	0.501064	0.593343	0.703536
finbertyiyang	0.793237	1.000000	0.450691	0.634784	0.755862
finvader	0.501064	0.450691	1.000000	0.310144	0.374579
gpt-3.5t	0.593343	0.634784	0.310144	1.000000	0.785985
gpt-4o	0.703536	0.755862	0.374579	0.785985	1.000000

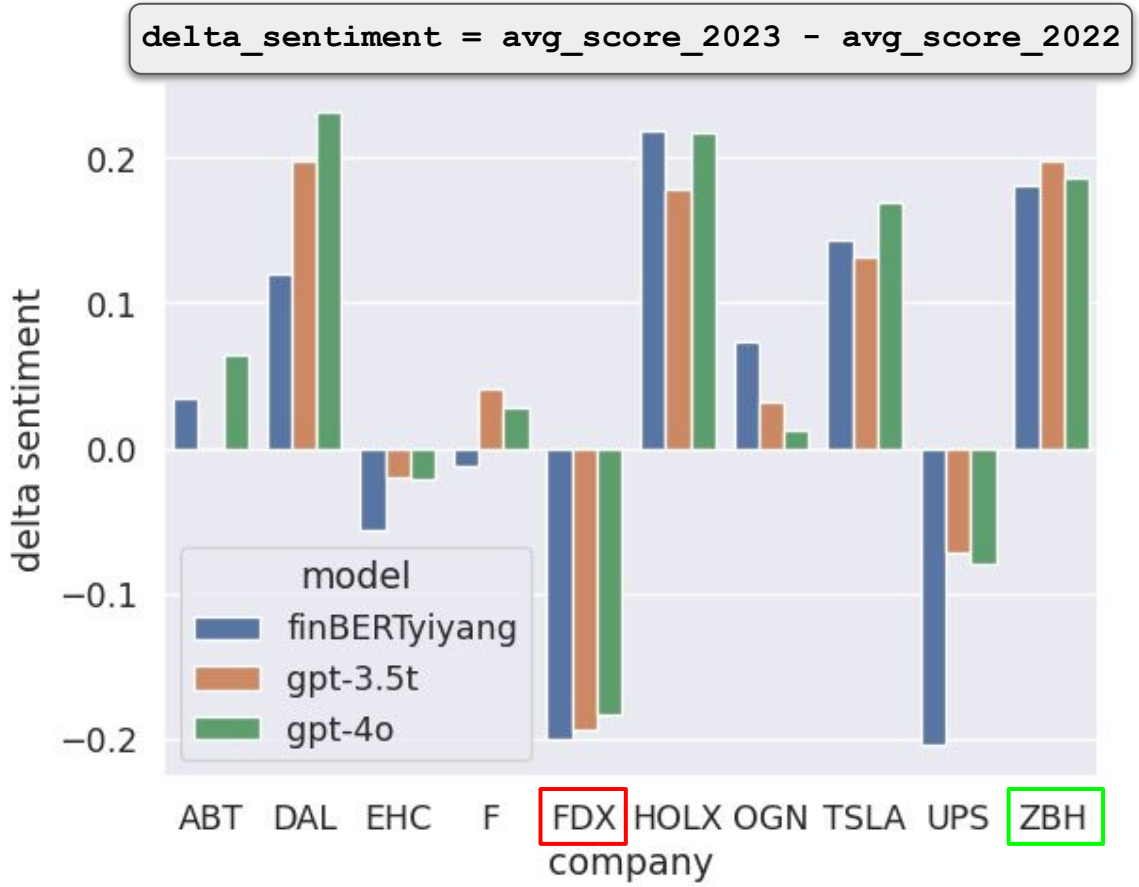
Results of finBERT and GPT show significant **correlation**

	finbertprosus	finbertyiyang	finvader	gpt-3.5t	gpt-4o
finbertprosus	1.000000	0.793237	0.501064	0.593343	0.703536
finbertyiyang	0.793237	1.000000	0.450691	0.634784	0.755862
finvader	0.501064	0.450691	1.000000	0.310144	0.374579
gpt-3.5t	0.593343	0.634784	0.310144	1.000000	0.785985
gpt-4o	0.703536	0.755862	0.374579	0.785985	1.000000

Results of finBERT and GPT show nice match



Results of finBERT and GPT show nice match



Highest delta: ZBH, +18.84%

Lowest delta: FDX, -19.20%

Zimmer Biomet Holdings Inc
NYSE: ZBH

Piaci összegzés > Zimmer Biomet Holdings Inc

FedEx
NYSE: FDX

Piaci összegzés > FedEx

RelevanceBert - Our own model

We wanted to replace GPT-4o in our pipeline, so we built our own model for selecting relevant sentences.

Simple supervised learning problem:

- ➔ Features - the sentences,
- ➔ Labels - the labels predicted by GPT-4o.

Surrogate model technique.

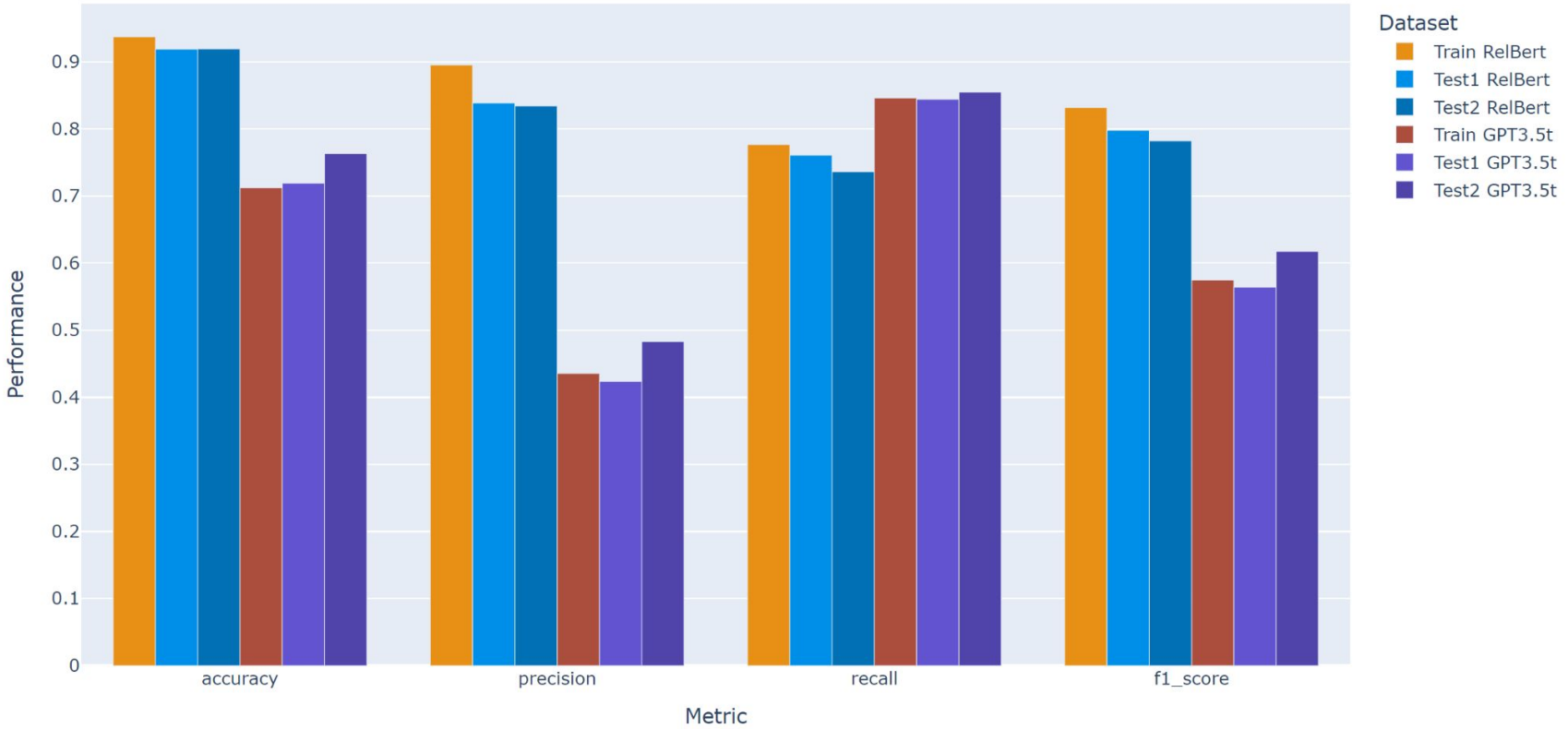
RelevanceBert - Our own model

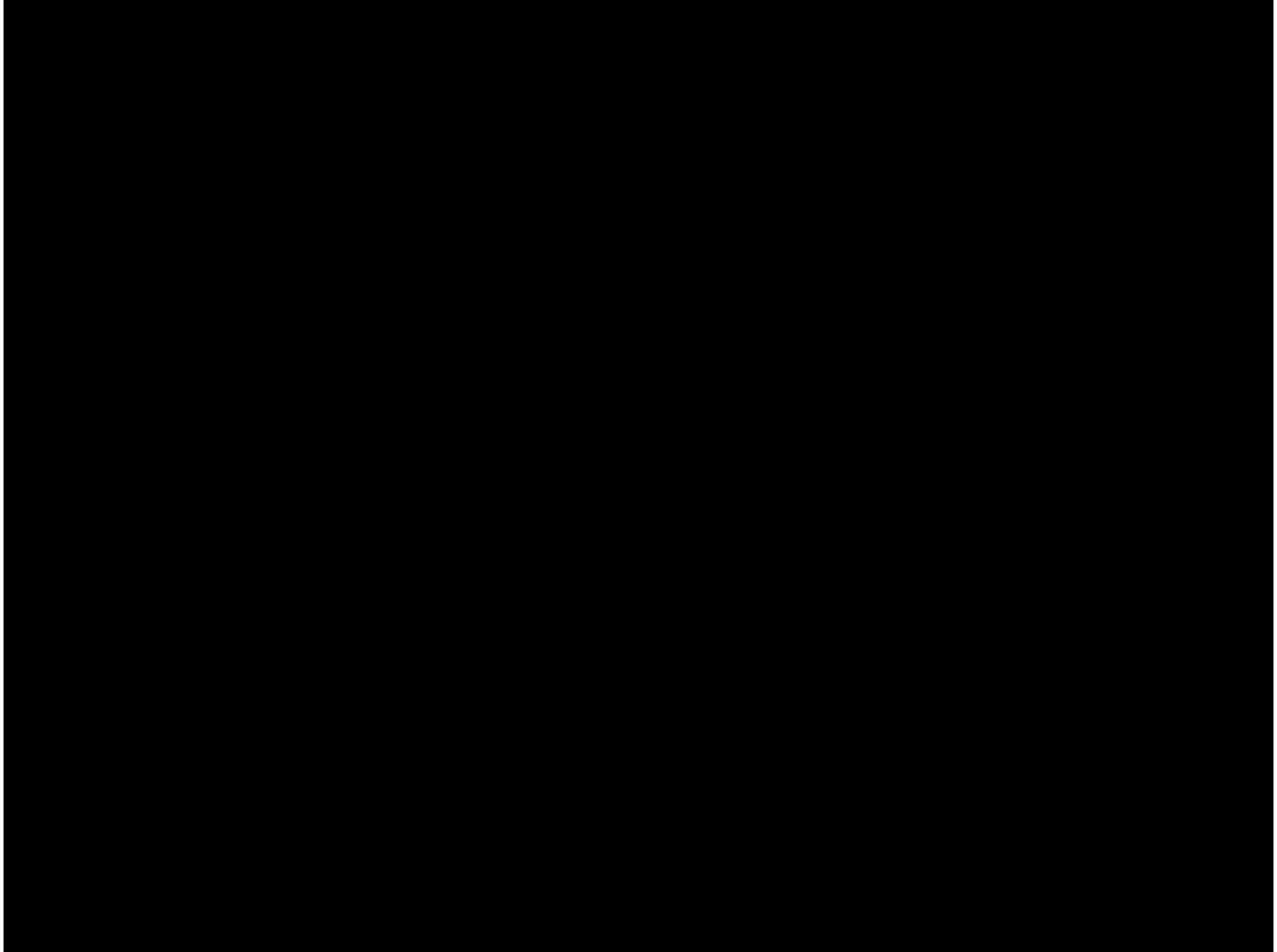
Fine-tuning the Finbert-Tone (Yiyanghkust) model via transfer learning.

- Dataset Descriptions:
  - **Train**: The training dataset contains 41,189 rows.
  - **Test1**: This test dataset has 4,357 rows, comprising reports from 2023 for Tesla and UPS. The training dataset includes their 2022 reports.
  - **Test2**: This test dataset includes 4,063 rows, with reports from 2022 and 2023 for Zimmer Biomet Holdings. The training dataset does not contain any data on ZBH.

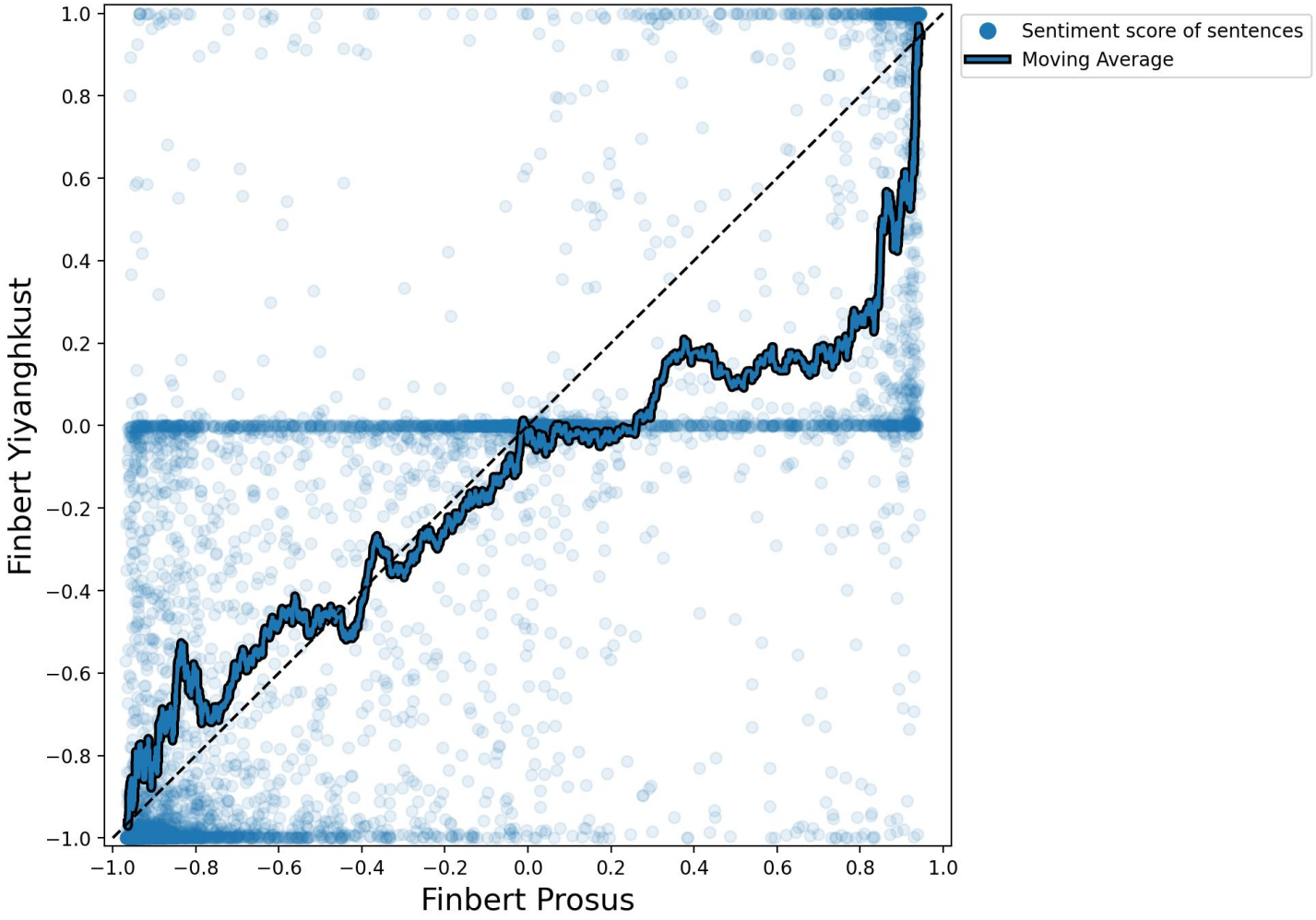
RelevanceBert - Our own model

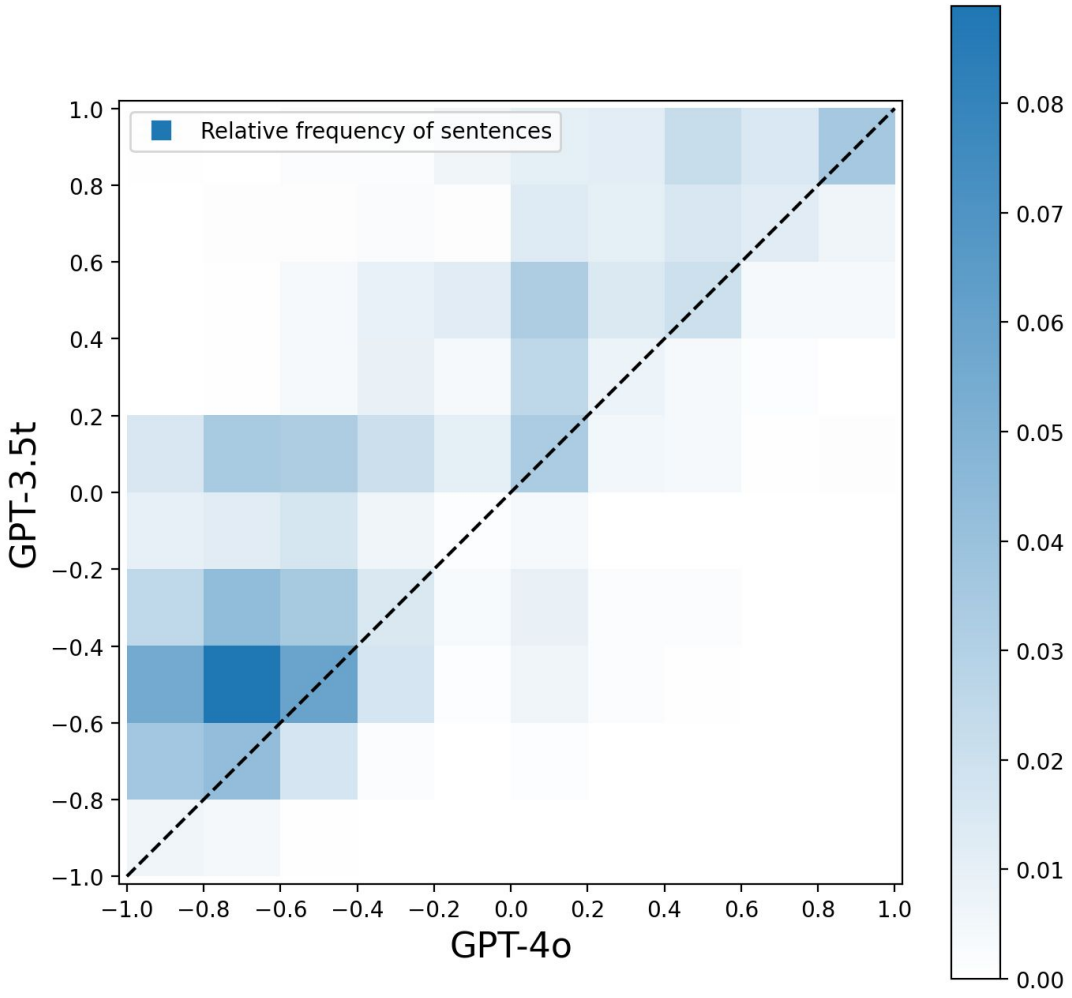
Model size
GPT-4o ~ 1500**b**
GPT-3 - 175**b** (turbo 20**b**?)
RelevanceBert - 100**m**

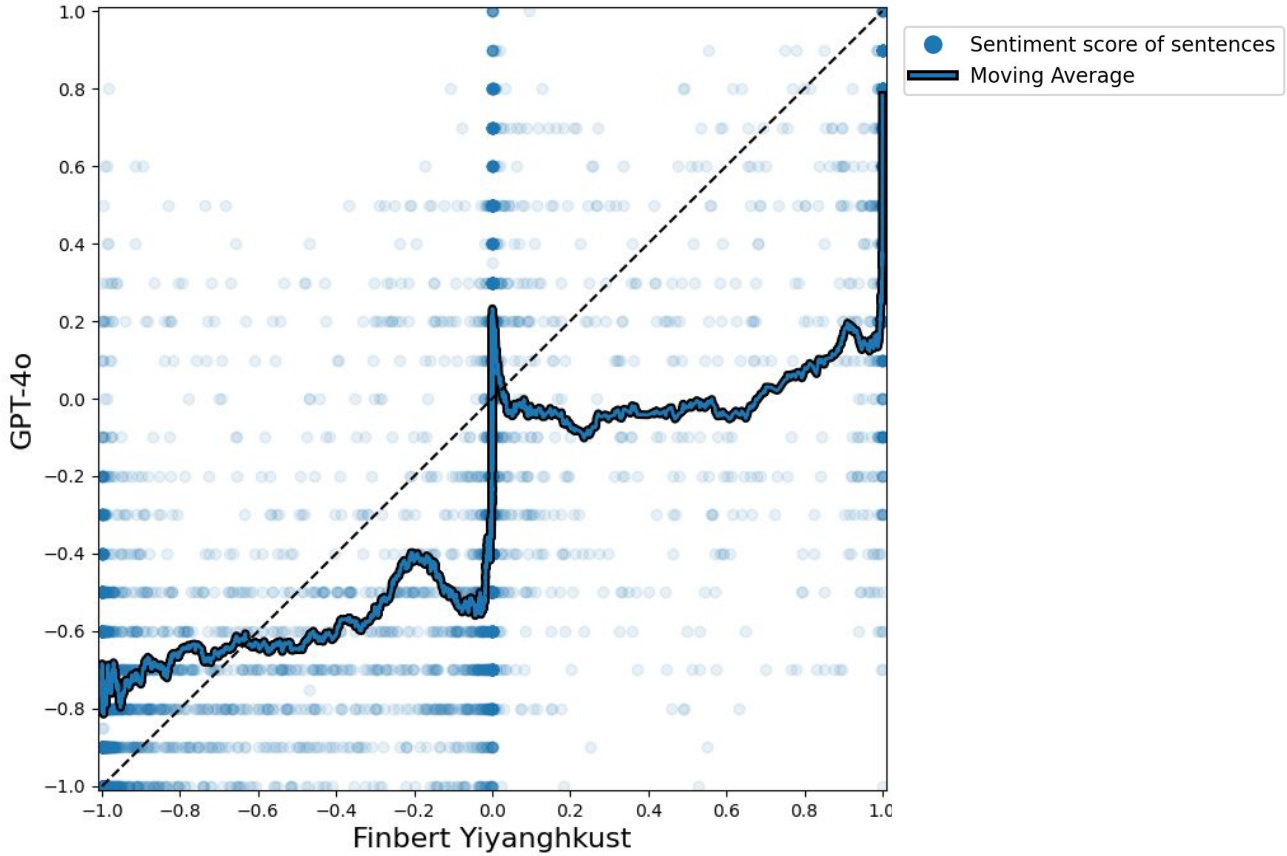




Supplementary slides


Sentiment score of sentences
Moving Average
Finbert Yiyanghkust
Finbert Prosus
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0


Relative frequency of sentences
GPT-3.5t
GPT-4o
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00


Sentiment score of sentences
Moving Average
GPT-4o
Finbert Yiyanghkust
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0