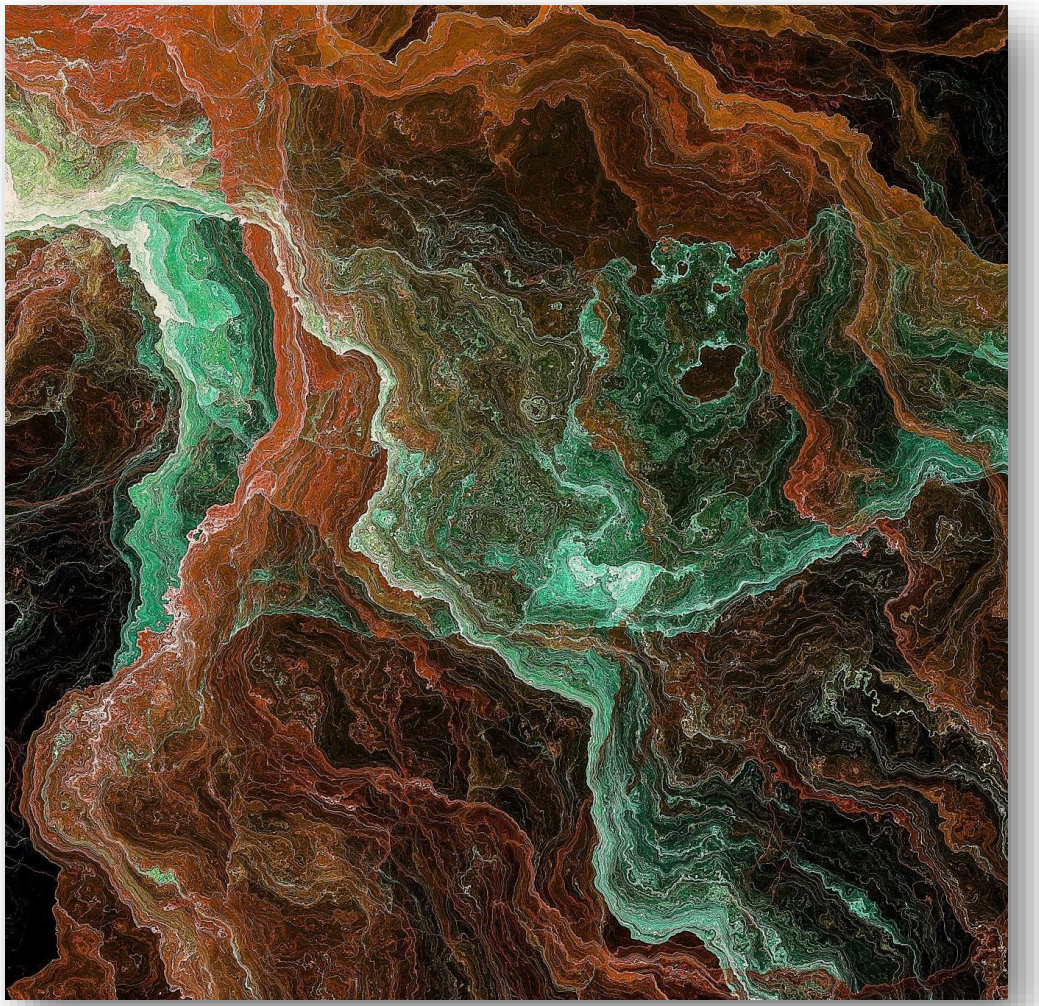


Unleashing the Potential of NLP: Exploring the Impacts of AI on Jobs and Tasks

MSCA 32018 IP01 (Spring 2023)
Natural Language Processing and Cognitive Computing

By

Kishor Kumar Reddy Mannur

Executive Summary with Meaningful Insights

- A refined dataset of about ~130k news articles was obtained from rigorous cleaning, preprocessing, and tokenizing of a massive dataset of ~200k news articles
- For a better understanding of the sentiment in the news articles' text, sentiment analysis was conducted to classify the articles into positives and negatives.
- To draw insights into the predicted sentiment texts, we apply **LDA** and **BERT** topic modeling techniques which helped us identify that the majority of the positive sentiments comprised of **health**, energy, efficiency, **market**, etc., as keywords within them hinting at a rise in the jobs and projects in these domains, for instance, healthcare, energy management, and others.

Executive Summary with Meaningful Insights

- However, the scenario on the other end of the spectrum is contrasting as we observed that the top failures and reasons for negative sentiments are **market**, **impact**, ai, **technology**, jobs, and **industry.**
- Overall, our study bridges the gap between complex AI research and practical knowledge, offering meaningful contributions to the advancement of the AI industry. By understanding the industry's growth trajectory and areas that require attention, we pave the way for informed decision-making, strategic planning, and the responsible utilization of this transformative technology

Actionable Recommendations

Some actionable insights that can be derived:

- **Capitalize on the Rising Trend of ChatGPT and OpenAI**: Given the surge in popularity and continued news coverage of ChatGPT and OpenAI, it is essential to explore opportunities in this domain.
- **Monitor Negative Sentiments for Risk Assessment**: The presence of negative sentiments surrounding topics like AI, technology, and jobs indicates potential risks and challenges.
- **Leverage Positive Sentiments in Health and Energy Domains**: The positive sentiments associated with keywords like health, energy, efficiency, and market indicate promising opportunities in these sectors.
- **Emphasize Human-AI Collaboration and Job Retention**: With the surge in AI capabilities, there is a concern about job displacement. Organizations should focus on promoting human-AI collaboration, reskilling initiatives, and creating new roles that leverage AI technology.
- **Foster Responsible AI Utilization**: As AI technology advances, it is crucial to emphasize responsible and ethical AI practices. strategies based on evolving industry dynamics.

Actionable Recommendations (in Project Working)

- Enhance data cleaning and preprocessing techniques to ensure dataset quality and consistency.
- Explore advanced sentiment analysis approaches for a more nuanced understanding of sentiment.
- Experiment with different topic modeling techniques to uncover additional latent topics.
- Validate results with domain experts to gain practical insights and interpretations.
- Implement interactive visualization tools for user-friendly exploration and presentation of results.
- Conduct longitudinal analysis to track sentiment trends and topic evolution over time.
- Publish and share findings to contribute to the knowledge base and foster collaboration.

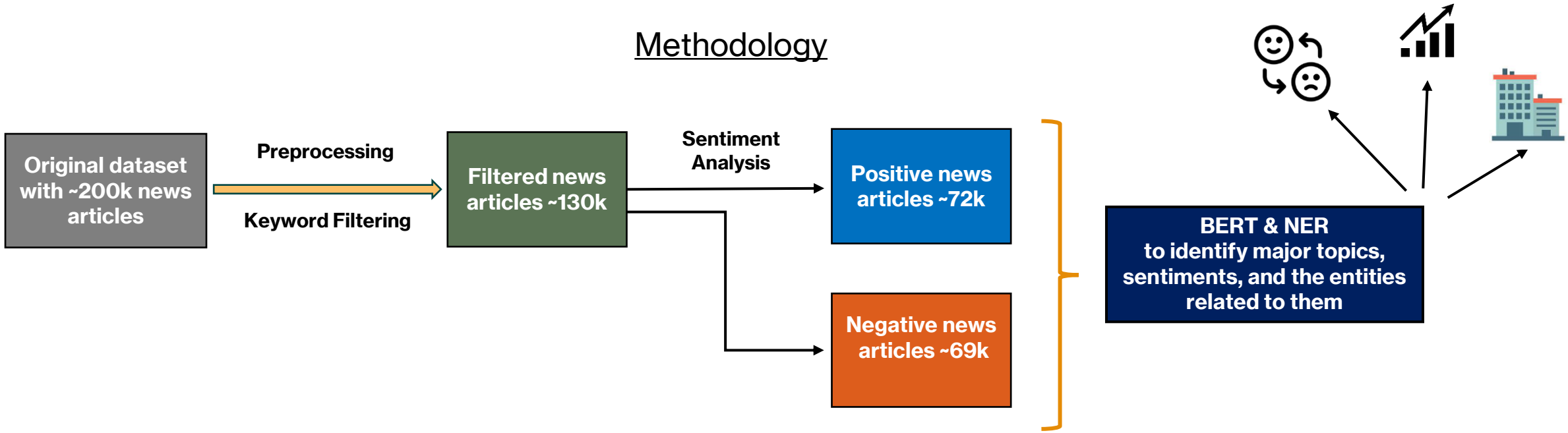
Article Clean-up and Filtering

The news articles datasets fetched using the starter notebook contain ~200,000 records.

After using the **filtration** and **cleaning** techniques **~134,000 records**

- Removed digits, non-word and non-space characters, newline, and tab characters
- Limited the word length for the article sentences &character length for the words inside the sentences
- Used closely related words along with the mentioned keywords to leave out unnecessary noise.

Methodology



Topic Detection & Sentiment Analysis Methodology

With the clean and filtered data at my disposal, the next aim is to:

- Perform sentiment analysis to identify the positively and negatively impacting news articles

  Using a **SVM** classifier, by training a custom Yelp **model**, I obtained the negative and positive sentiments

- Further, identifying the sentiment changing over time and what is causing that. (topic modeling for positive and negative news articles)

  By applying **BERTopic** on both the categories of sentiment-analyzed news articles, **tuning** the parameters like the number of topics, etc. I found the major topics causing the sentiment.

  Although LDA is good for visualizations, BERT takes it up a notch by giving the top words in the topic making it **more interpretable and user-friendly** which can help in conjunction with NER

- Eventually, I would apply NER to these sentiment-classified articles to obtain information on who is responsible for the trends and where exactly are they happening

Topic Modeling

BERTopic applied to positive and negative sentiment articles gives us a better understanding of the trends.

- Removed stopwords
- Lowered the case of the text

– Starting with **n=100** topics and moving it down all the way to **n=10** to obtain precise results.

Positive Sentiment Articles Topics

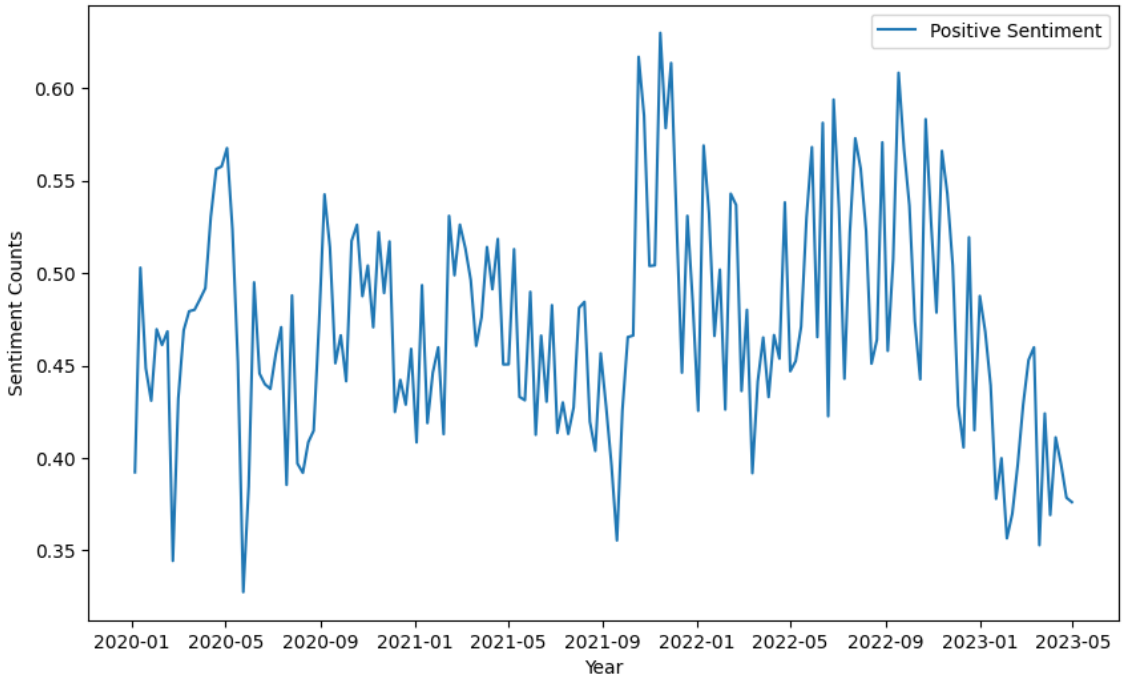
	Topic	Count	Name
0	-1	30709	-1_ai_data_gray_media
1	0	5088	0_market_healthcare_artificial_intelligence
2	1	2081	1_ai_data_learning_machine
3	2	1584	2_energy_data_solar_new
4	3	1533	3_nvidia_edge_ai_performance
5	4	1452	4_market_report_artificial intelligence_artifi...
6	5	1223	5_market_machine learning_machine_learning
7	6	875	6_art_images_image_artists
8	7	665	7_chatgpt_gpt_openai_chatbot
9	8	616	8_conversational_voice_conversational ai_customer

Negative Sentiment Articles Topics

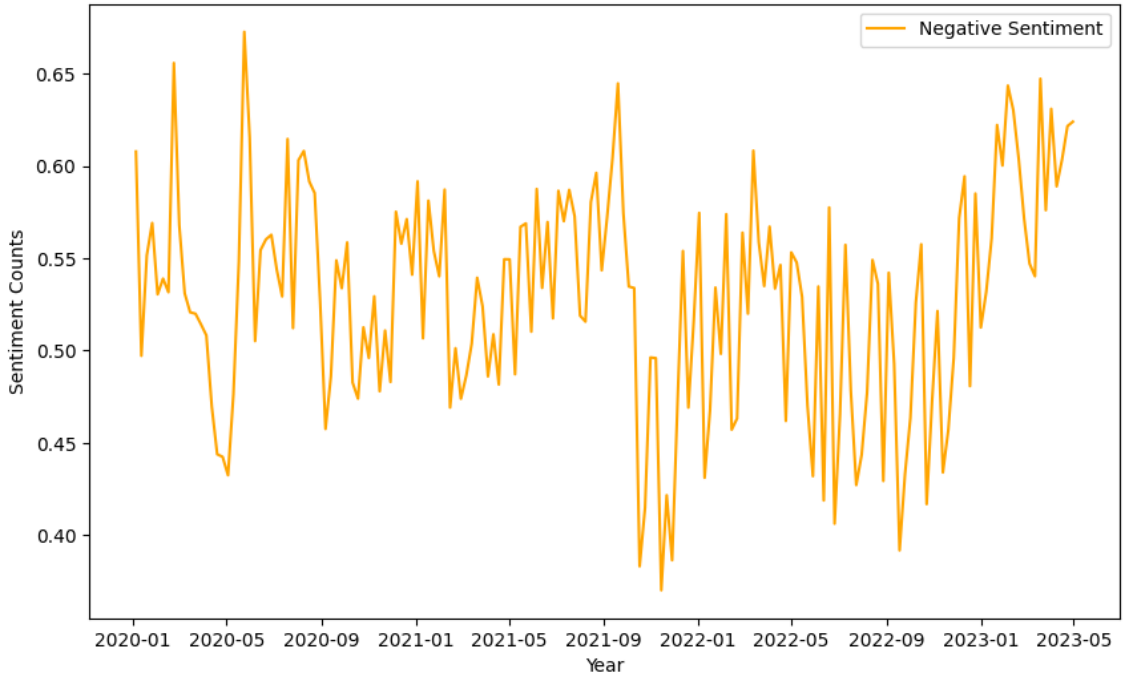
	Topic	Count	Name
0	-1	33557	-1_ai_data_gray_media
1	0	1661	0_energy_new_solar_data
2	1	1568	1_ai_data_learning_machine
3	2	1471	2_nvidia_edge_ai_performance
4	3	1353	3_market_machine learning_machine_learning
5	4	1160	4_market_artificial intelligence_artificial_in...
6	5	893	5_market_healthcare_intelligence healthcare_ar...
7	6	828	6_art_images_artists_dalle
8	7	723	7_chatgpt_gpt_openai_chatbot
9	8	641	8_healthcare_covid_medical_patients

Sentiment Analysis Overall

Avg Positive Sentiment Score vs Time



Avg Negative Sentiment Score vs Time



Sentiment Scores (Positive & Negative) over Time

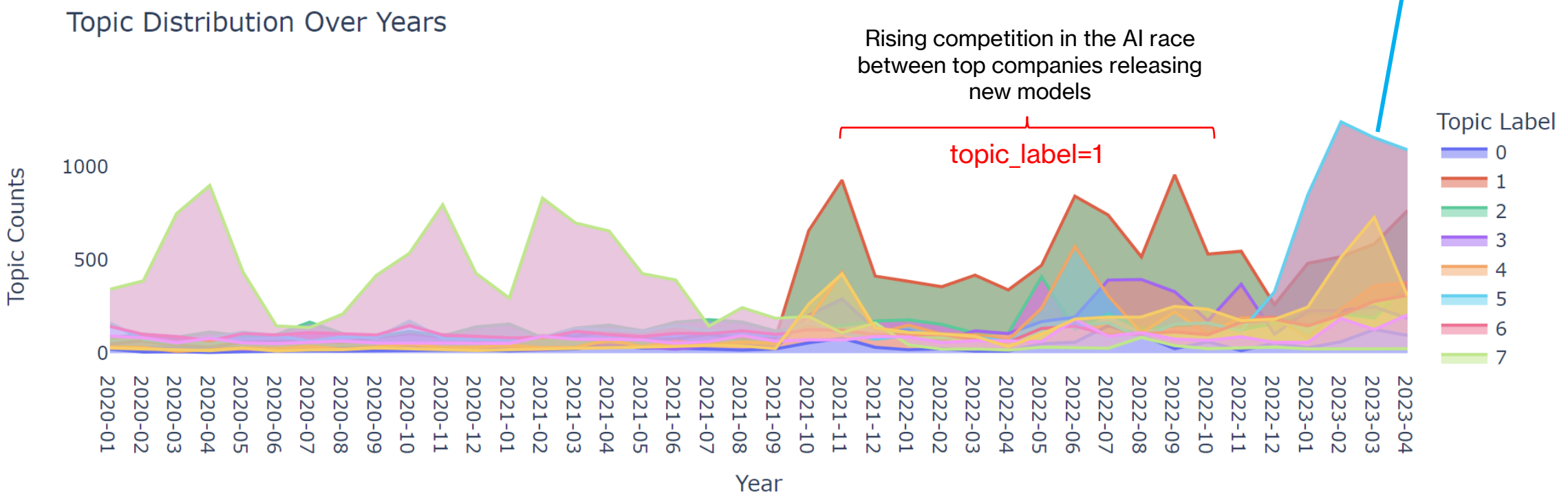
Topic 4: Market, report, artificial intelligence, global, industry, research, growth.

Topic 5: ChatGPT, AI, OpenAI, Google, search, Microsoft, chatbot, technology.

Sentiment Analysis (Positive)

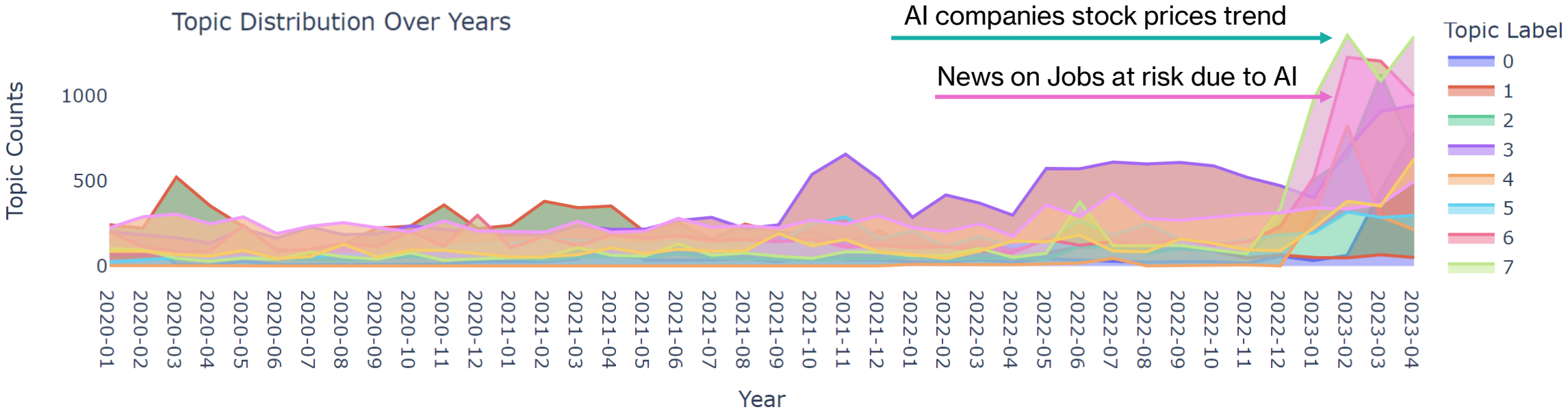
- We can see the progression of the top topics over the years
- The **topic_label=5** has surged since late 2022 and still making news, interesting, right? It's about showing the LLM introduction and uptrend, especially in the early 2023s with GPT 3.5 out.
- We can also see the topic stays in trend for about a year showing the race of AI in the year 2022

ChatGPT & OpenAI



Sentiment Analysis (Negative)

- It is interesting to see that negative sentiment isn't particularly low and can help identify infamous topics/ trends.
- **topic_label=7** rose in mid-2022 and gained momentum towards the end of 2022 which shows how the stock prices of AI companies dropped, however, it suggests that due to **Google's loss of $100B in shares, thanks to Bard's introduction video blunder**, making it a trend.
- **topic_label=6** discusses AI models and systems, their capabilities, the role of human interaction, and the impact of AI technology on various domains in 2023. With the surge in AI and its capabilities, humans are at risk of losing their jobs.



Entity Identification

- It is important to extract the named entities from the news articles which can lead to actionable insights. Using **NER with Spacy**, we extract the entities like **Person**, **Organization**, and **Location** among others.
- We obtain the following entities from both positive and negative news articles respectively as follows:
- By leveraging the topic modeling, sentiment analysis from earlier and using the NER results, we can perform a targeted sentiment analysis and gain information and insights

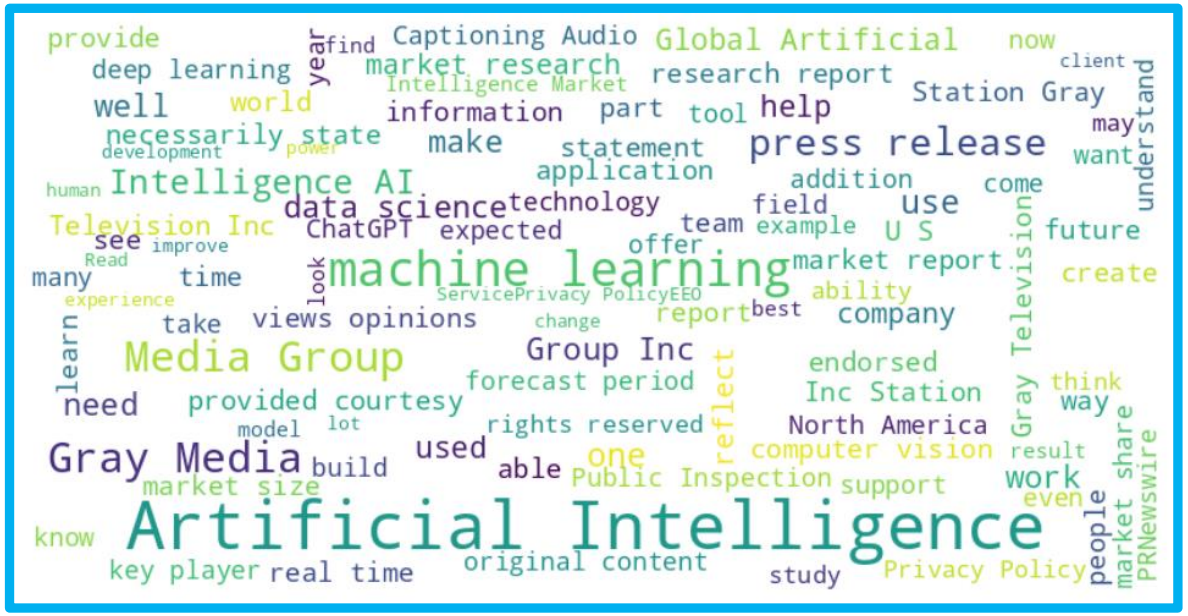
Positive Sentiment Entities

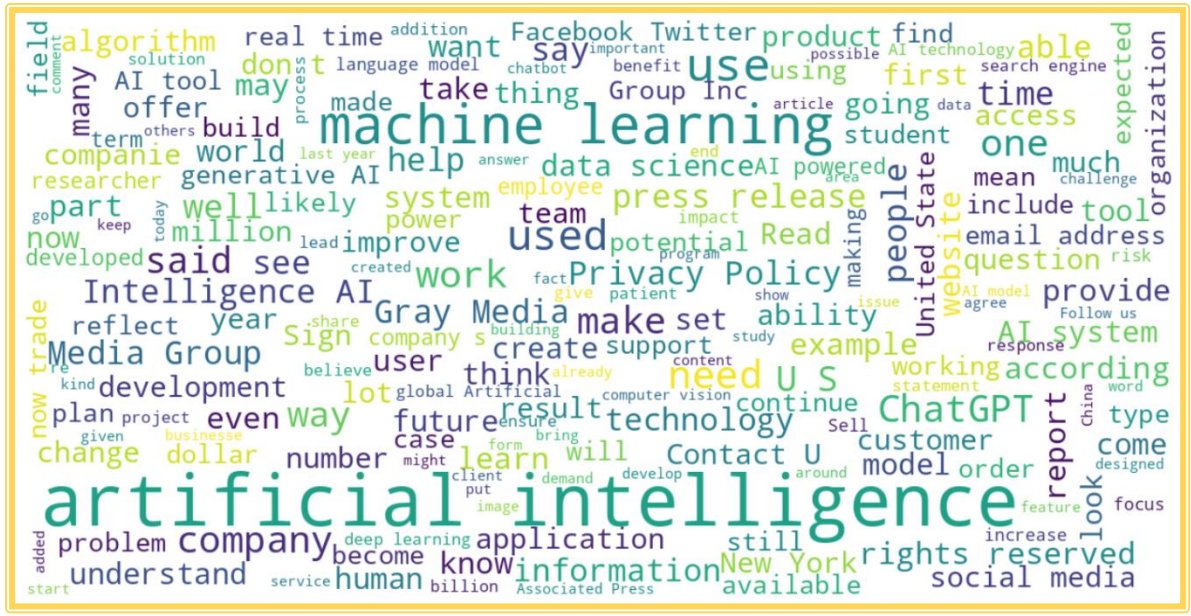
```
[('microsoft', 19651),
 ('google', 15940),
 ('ibm', 15543),
 ('media group, inc.', 15510),
 ('europe', 14772),
 ('north america', 12369),
 ('intel', 7886),
 ('africa', 5531),
 (', inc.', 3820),
 ('asia pacific', 3426),
 ('dr.', 3376),
 ('quantum', 3273),
 ('middle east', 3154),
 ('sony', 3135),
 ('nvidia', 3059)]
```

Negative Sentiment Entities

```
[('microsoft', 45112),
 ('google', 44470),
 ('ibm', 10560),
 ('europe', 9772),
 ('media group, inc.', 6908),
 ('npr', 6019),
 ('intel', 5511),
 ('north america', 5081),
 ('fcc', 5076),
 ('gpt', 4722),
 ('pichai', 3652),
 ('africa', 3542),
 ('cnn', 3425),
 ('nvidia', 3248),
 ('fda', 3175)]
```

Entity Identification Word Clouds

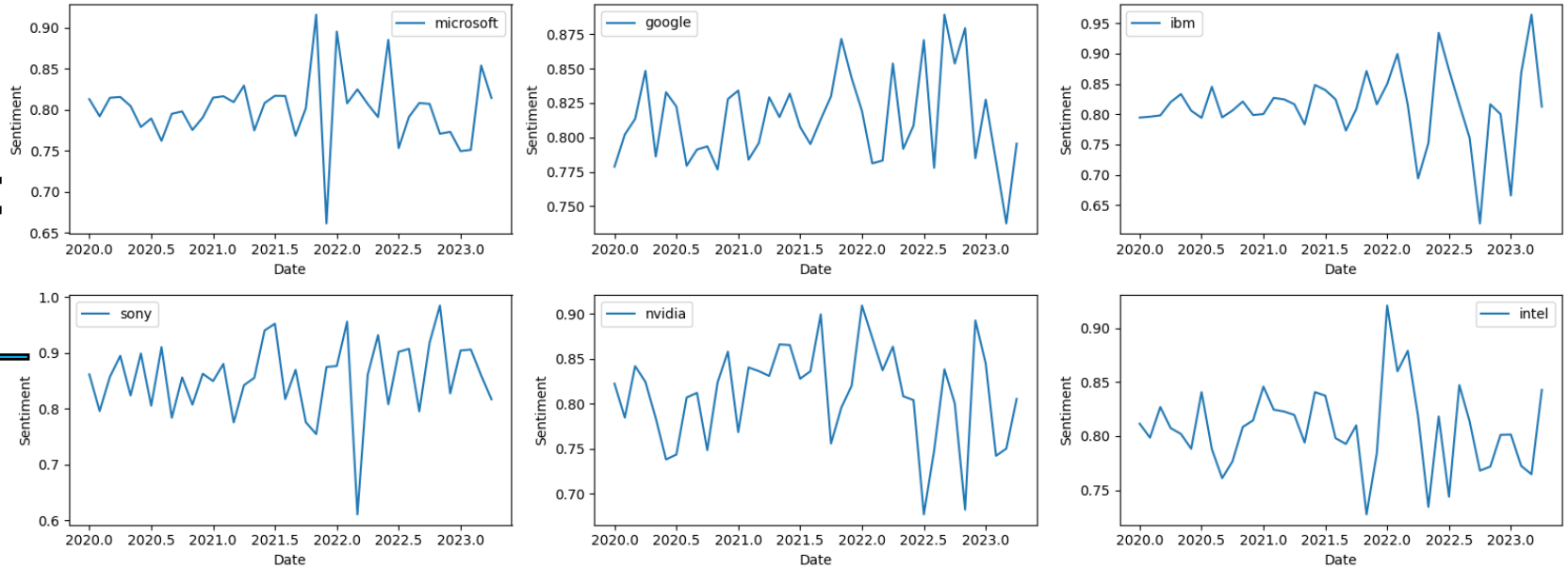


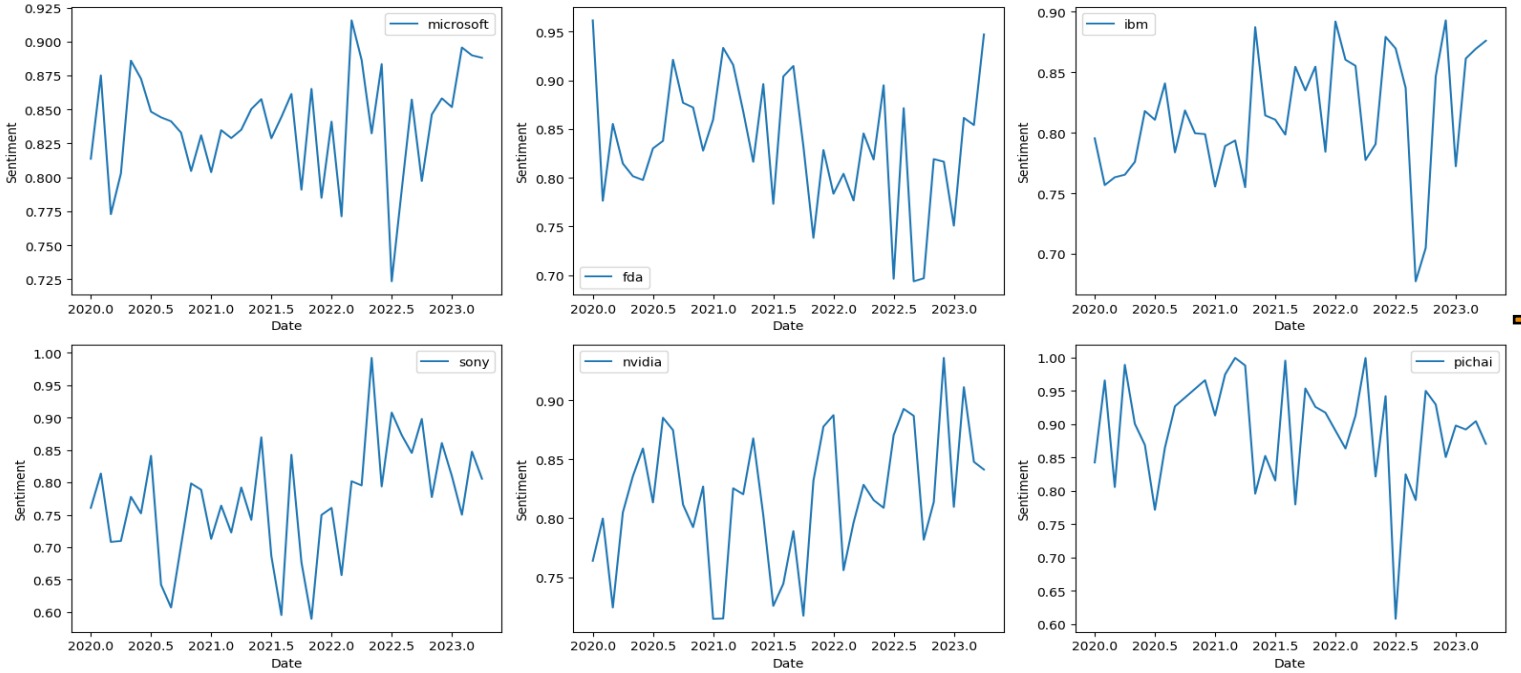


- From the extracted positive word cloud, it is suggestive of the topics in **press**, release, information, **original content**, **improve,** etc.
- The further targeted sentiment can be observed that there was an overall rise in sentiment from mid ‘2022 signaling at the release and news surrounding ChatGPT
- On the other hand, the negative word cloud shows terms like people, **privacy**, policy, customer, **rights**, **problem**, etc., which show the privacy concerns revolving around the advancements in the field of AI and machine learning. Microsoft sentiment both positive and negative rose in late ‘22.

Entity Identification
Targeted Sentiment

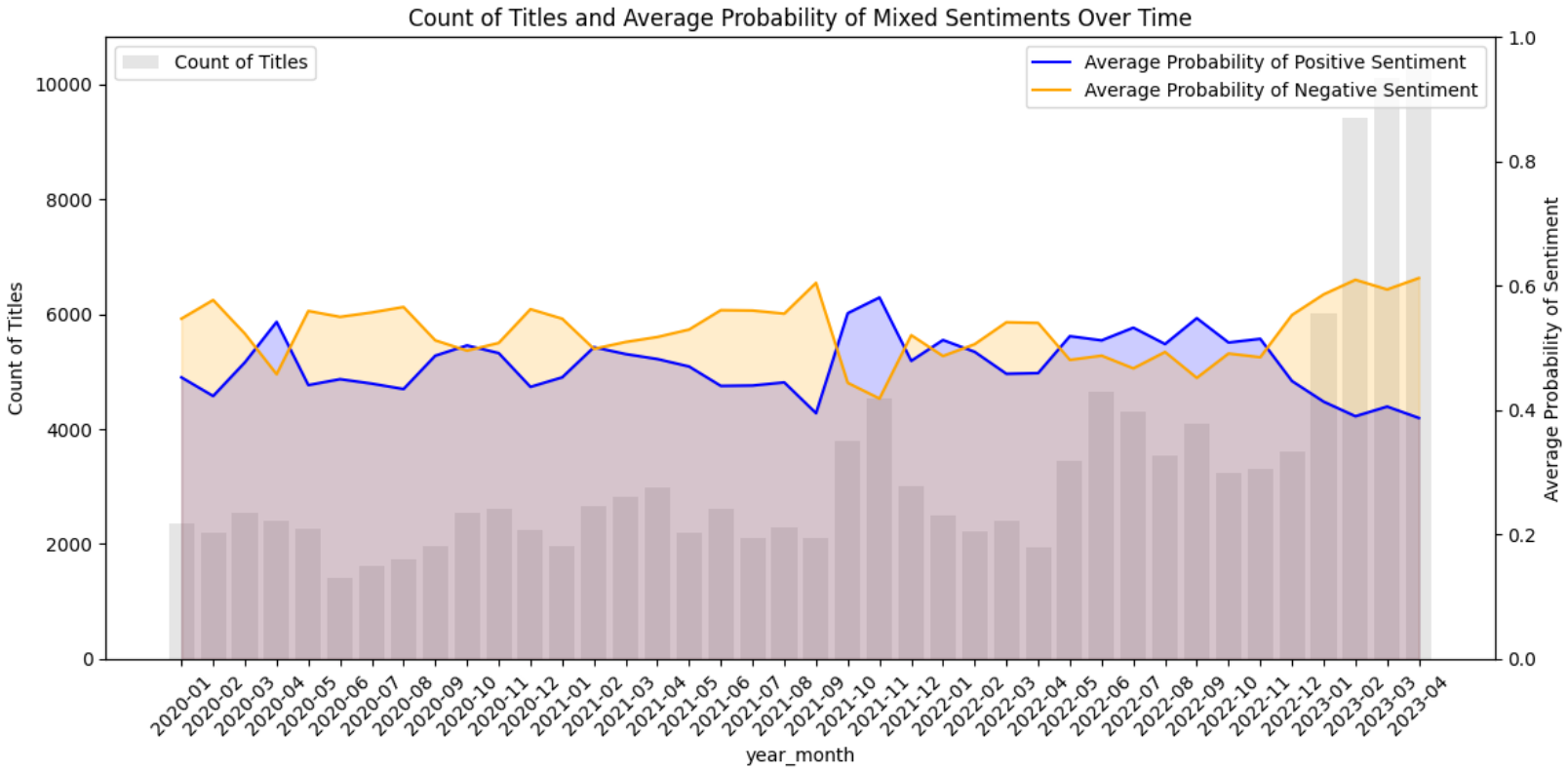
Targeted Sentiment Analysis on the Positive NER obtained





Targeted Sentiment Analysis on the Negative NER obtained

Appendix



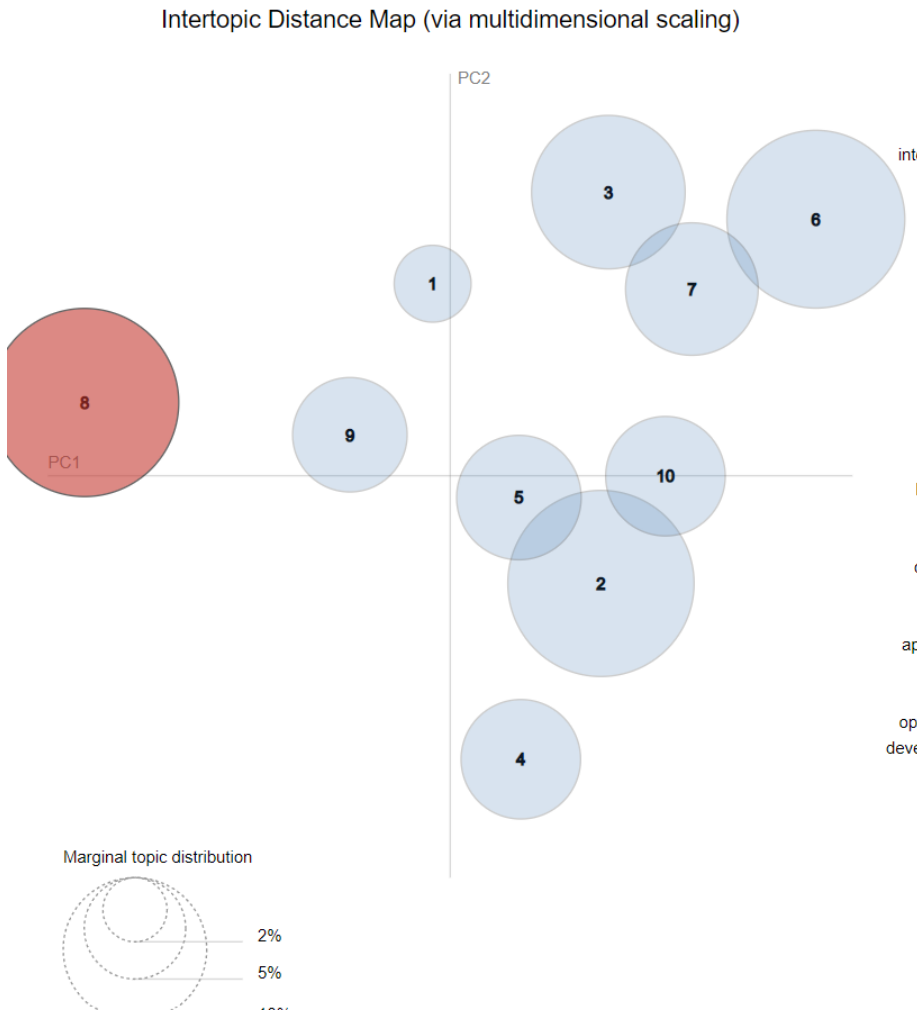
Sentiment Analysis showing mixed opinions over time

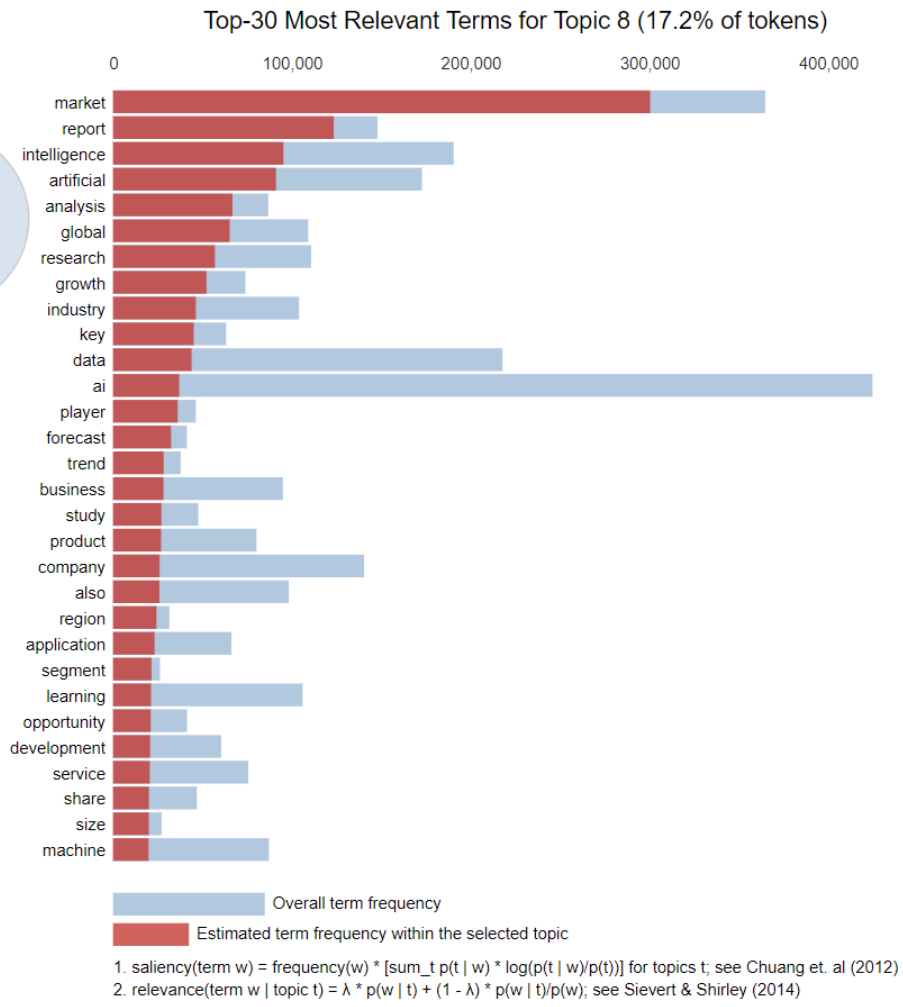
This graph shows us the general mixed sentiment against the count of articles

ChatGPT and OpenAI find themselves having the most **positive** sentiment and trending upwards

Humans are at **risk of losing their jobs.** Unstable **stock prices** in the AI arena.

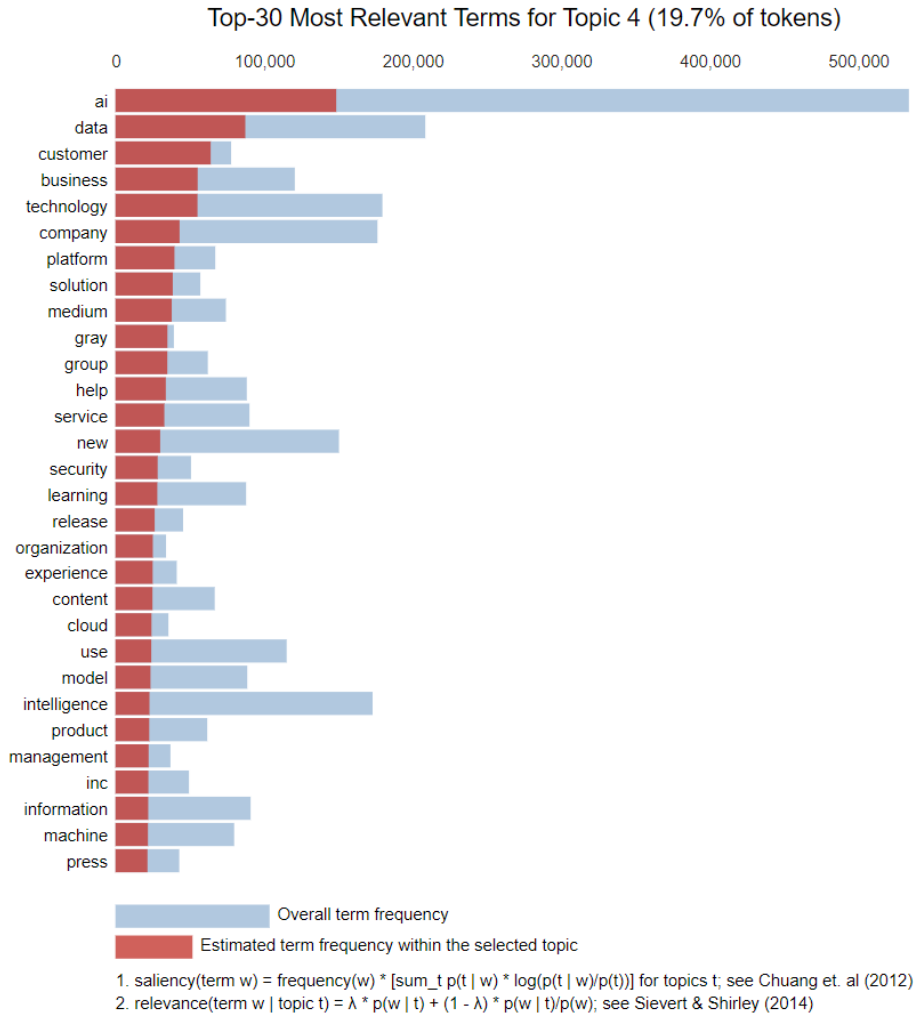
Appendix





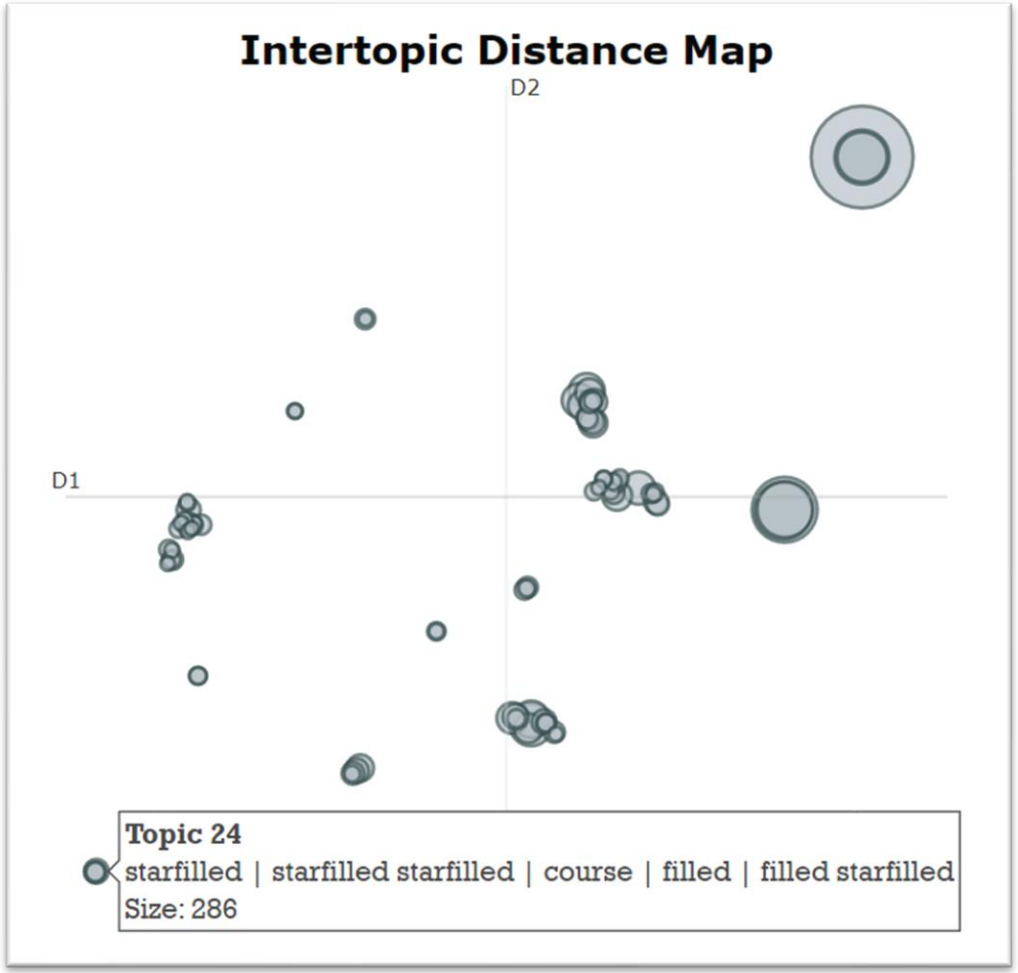
Positive Sentiment pyLDAViz for Topic Modelling

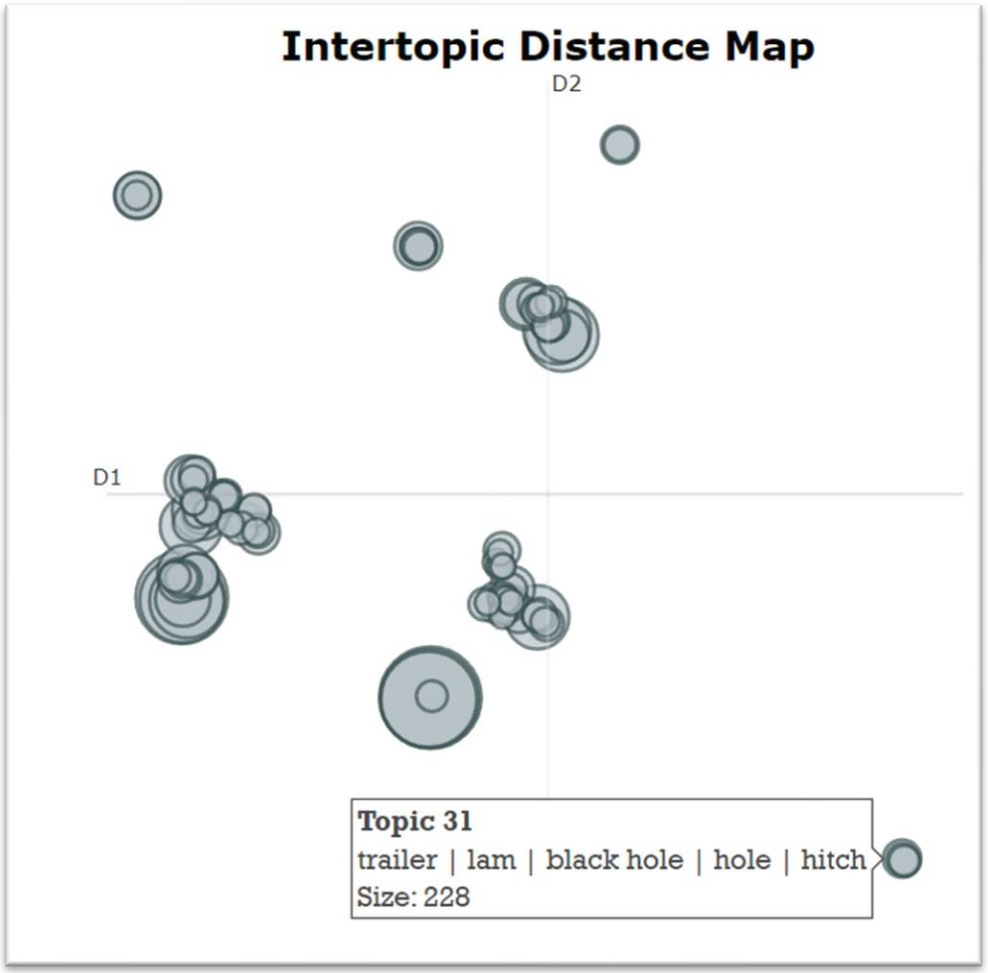
Appendix



Negative Sentiment pyLDAViz for Topic Modelling

Appendix





Positive and Negative Sentiment BERT Topics Visualizations