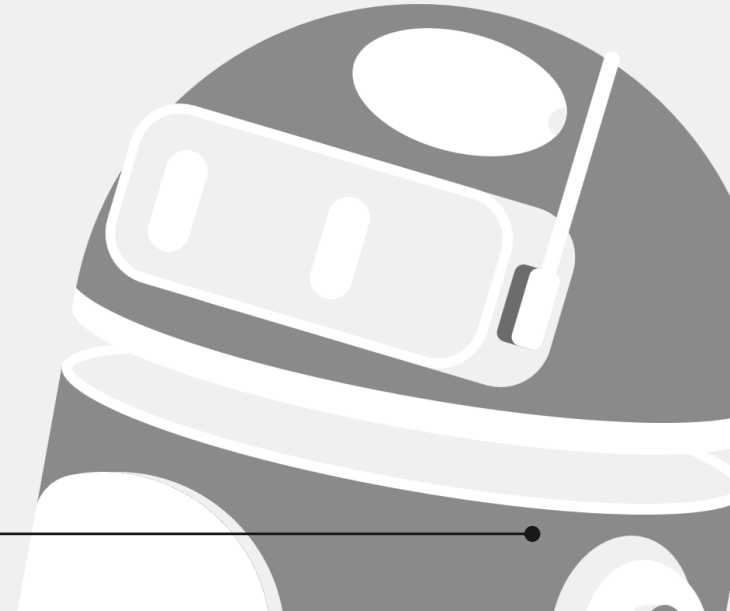
AI Impact Insights: Analyzing Trends in Jobs and Industries



By Sai Krishna Jasti

Agenda

01 → Executive Summary and Meaningful Insights

02 → Actionable Recommendations

03 → Article Clean-up and Filtering

04 → Sentiment Analysis

05 → Sentiment Over Time – Analysis and Visualization

06 → Topic Detection

07 → Entity Identification

08 → Targeted Sentiment Identification

Executive Summary

What did I do?

- ✓ Collected and analyzed a collection of news articles on Data Science, Machine Learning, and Artificial Intelligence.
- ✓ Filtered and cleaned the articles to eliminate noise and focus on relevant information.
- ✓ Applied sentiment analysis to understand the overall sentiment and trends within the articles.
- ✓ Utilized entity analysis to identify the sentiments associated with different entities mentioned in the articles.

How did I do?

- ✓ Implemented a rigorous data cleaning and filtering process to ensure the data is relevant and of good quality.
- ✓ Used a range of techniques for sentiment analysis to **extract sentiment scores** and understand the overall sentiment of the articles.
- ✓ **Identified the positive and negative topics** discussed in articles using the Gensim LDA model.
- ✓ Leveraged spaCy's entity recognition models to identify and **track sentiment changes for specific entities** mentioned in the articles.

What did I achieve?

- ✓ **Identified major topic keywords** such as Market Intelligence, Research, Risks, and Opportunities from the analyzed articles.
- ✓ Revealed insights into the **impact of AI** on different industries and job lines, highlighting areas with high potential for automation and AI adoption.
- ✓ Provided **actionable recommendations** based on the analysis, helping organizations understand the implications of AI and suggesting ways to leverage it for automation and improved productivity.

Meaningful Insights

Impact of COVID-19 on AI Sentiments:

- During Q1 2020, there was a notable **decline in sentiment**, likely **influenced by the COVID-19** pandemic.
- Negative sentiments could be due to the concerns regarding **AI's use for tracking people's movements** during lockdowns.
- The sentiment **gradually recovered**, indicating a shift in priorities and perception.

Positive Topics and Trends:

- Positive sentiments revealed **promising growth opportunities in various industries leveraging AI**.
- Businesses explored **AI to drive growth, implement client-centric services, and automate processes**.
- Key topics included the **application of AI** in manufacturing and consumer goods, emerging tools like ChatGPT, and the competitive landscape among major players like Google and OpenAI.

Concerns and Risks:

- Negative sentiments highlighted concerns related to **data privacy, risks** associated with AI and automation, and potential **weaponization of AI technologies**.
- Articles showed that opinions varied on AI's role in replacing jobs, and **ethical considerations** gained attention in discussions.
- **Implementation of GDPR** on AI in 2022 also had a setback on sentiments, because of **its challenges in adaptation**.

Entity Level Sentiment Analysis:

- Visualize sentiment variations over time for key entities discussed in positive and negative sentiment articles.
- This analysis provides insights into the changing perception and sentiment towards these specific entities.

Actionable Recommendations

Enhanced Infrastructure and Quality

- Invest in scalable cloud platforms and efficient data pipelines for seamless data flow.
- Enhance data quality through data cleansing tools and data governance policies.

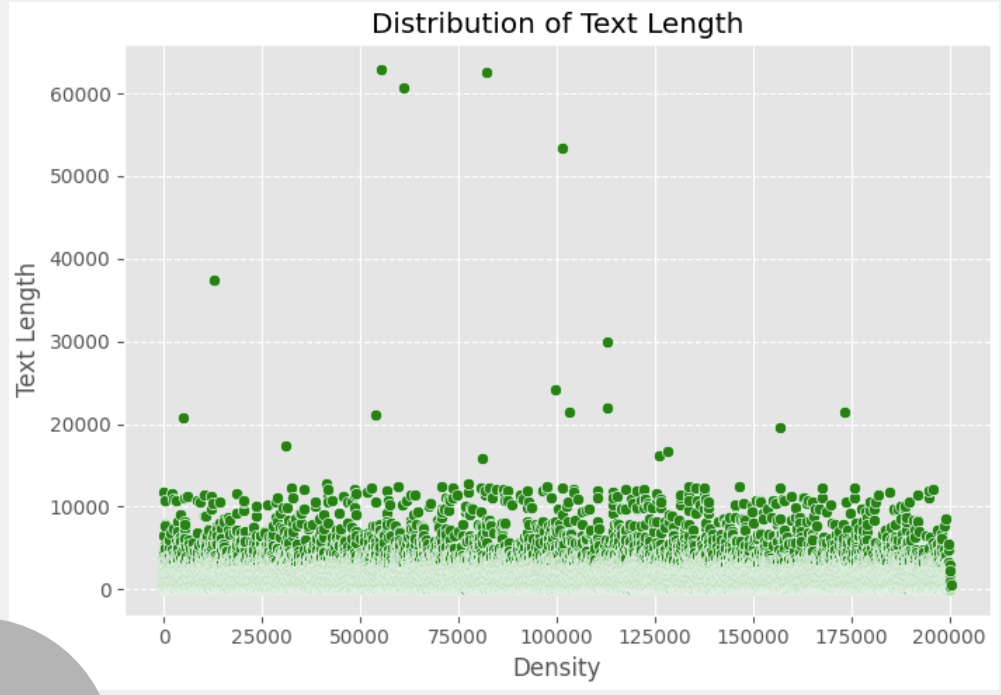
Data Governance and Ethical AI

- Define data ownership, enforce quality standards, and ensure compliance with privacy regulations.
- Establish clear ethical guidelines for AI development and deployment.

Foster Collaboration

- Encourage collaboration between academia, industry, and government entities.
- Promote knowledge sharing and research partnerships to advance AI technologies.

Article Clean-up and Filtering



Filtering:

- The original data had 200K records out of which 20K records were discarded as irrelevant articles.
- This was done by dropping the duplicates and handling outliers based on the length of the articles (percentiles).

Status	Record Count
Original Dataset	200K records
Duplicate Removal	-2K records
Text Length based clean up	-18K records
After Filtering	180K records

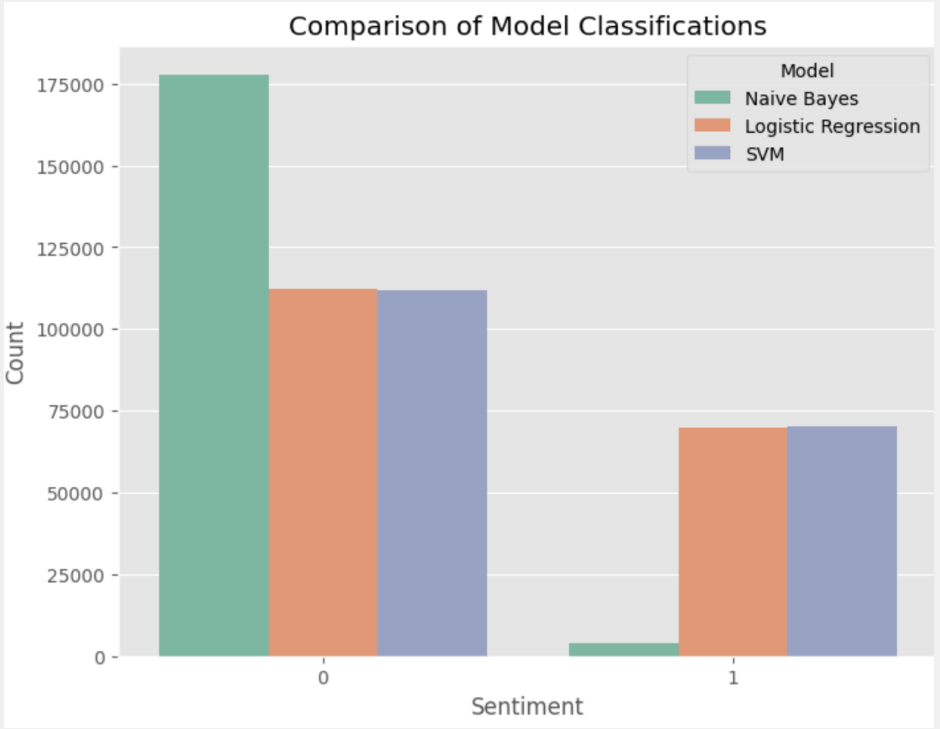
Article Clean-up and Filtering

Cleaning:

- The date column has been properly formatted, **extracted the year, quarters, months and day of the week from date**.
- Title and Text columns have been cleaned of **remnant web crawls, non-ASII characters, punctuations, extra spaces, new line characters, non printable characters and stop words**.
- **Regex and Beautiful Soup** were used to perform this cleanup.
- **The cleaned columns are then tokenized** using NLTK's word tokenizer for further analysis.

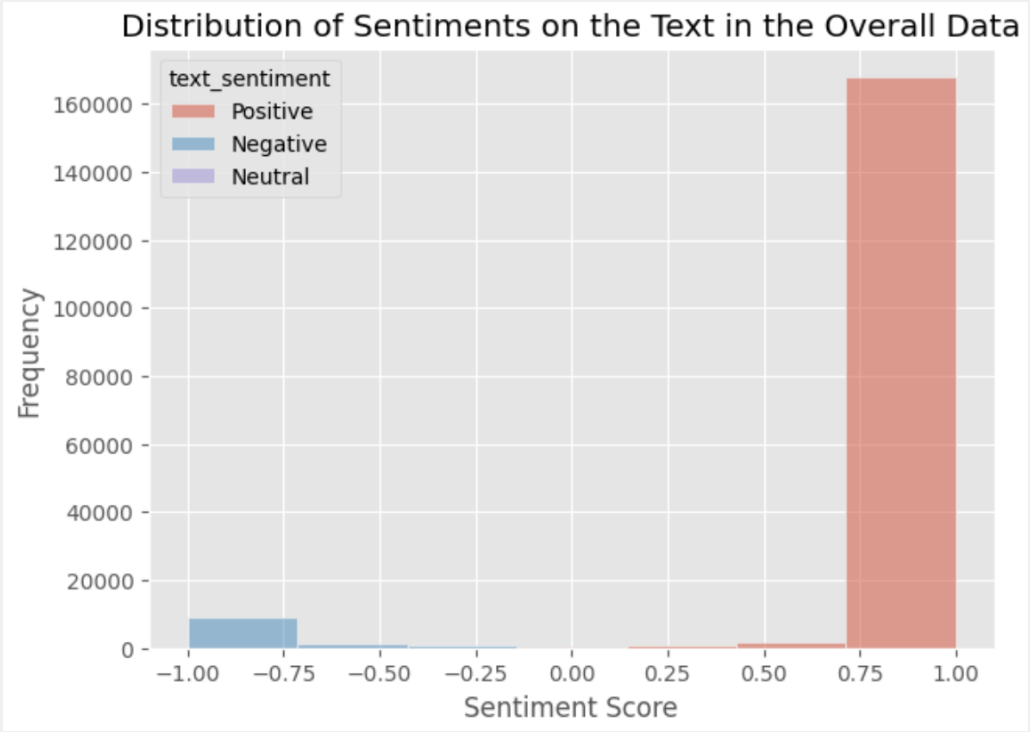
Sentiment Analysis

- Sentiment Analysis was initially done using **custom pipeline built using Yelp Dataset.**
- Three pipelines were built using Naïve Bayes, Logistic Regression and SVM. **Logistic Regression was performing better of the three**.
- Both the Logistic Regression and SVM were able to achieve an **accuracy of >95%** on the test sets.
- The sentiments as classified into **Positive (1) or Negative (0)** using this.
- Some of the results were verified and it **was found that the model failed to classify several articles as Positive even though they are**.
- The results were not quite satisfactory since the data that was used to train the **custom pipeline is Yelp reviews while we are trying to classify the sentiment on News Articles**.



Sentiment Analysis

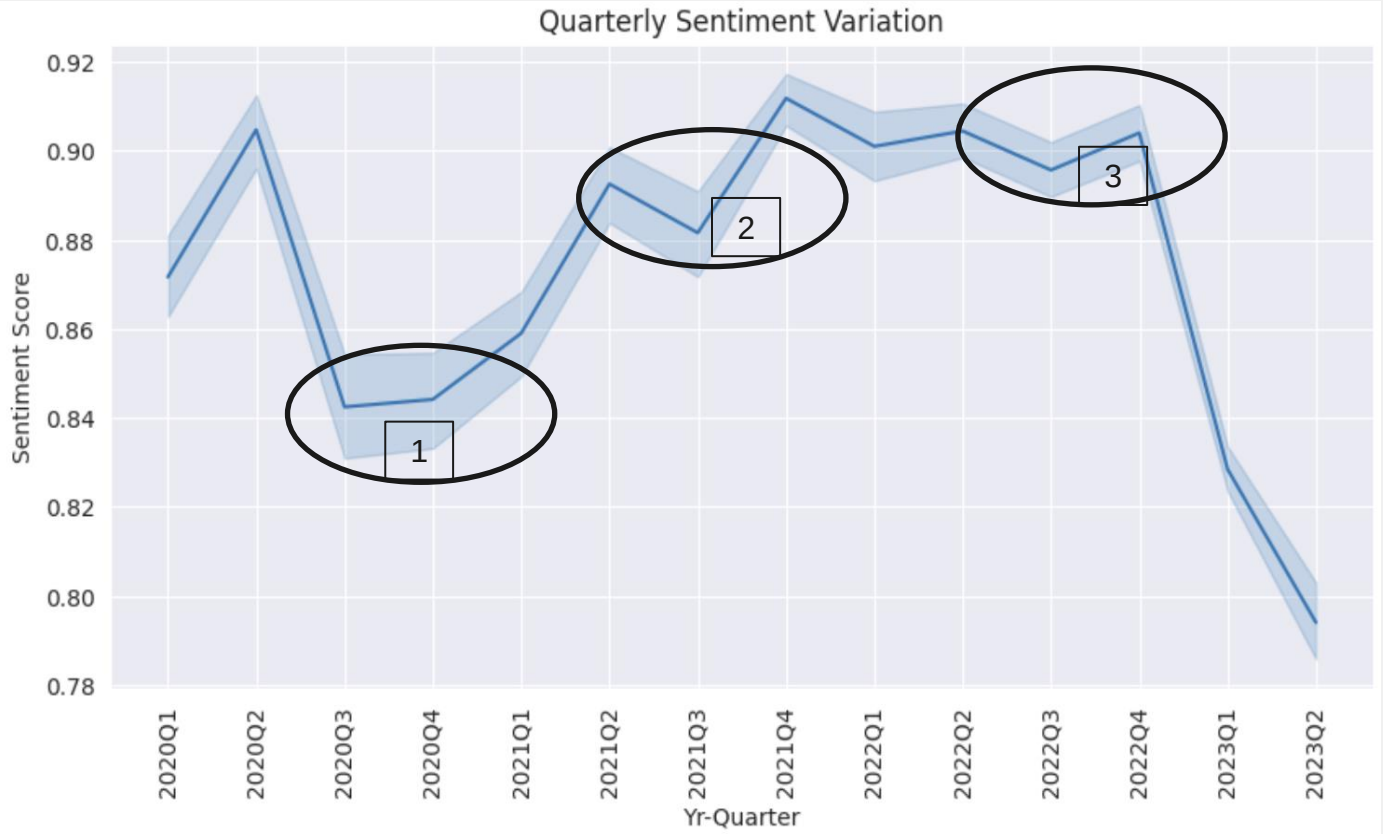
- To achieve better results, VADER was considered for sentiment classification. **VADER is a lexicon and rule-based sentiment analysis tool.**
- **VADER is capable of understanding and interpreting the contextual valence of words and phrases**, which is essential for sentiment analysis.
- The **discrepancies** observed with custom model have been greatly **nullified with the dictionary-based VADER** approach.
- Out of the 180k articles, **~168k were positive and the rest ~12k** are negative in sentiment about Data Science, Machine Learning and AI.



Sentiment Over Time – Analysis and Visualization

Overall Sentiments:

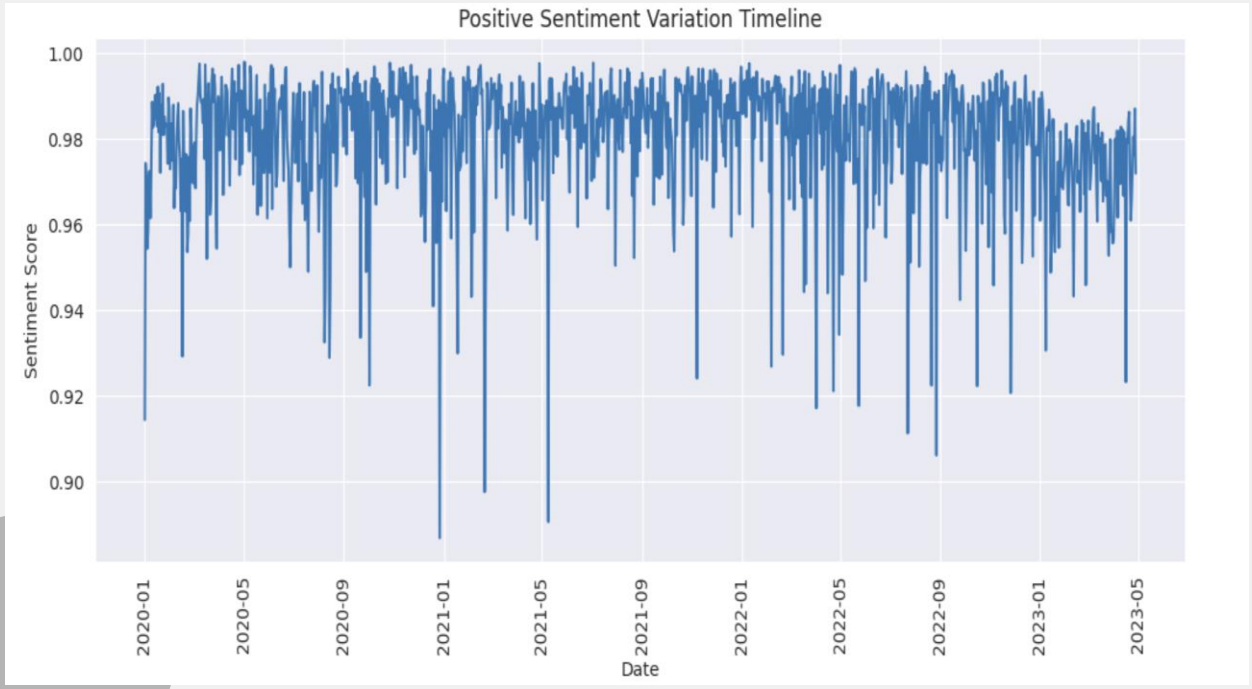
- From the overall sentiments, three major variations can be observed.
- First one was during **the COVID-19 Pandemic** where the sentiment greatly diminished as people believed **AI could be tracking their movements in lockdown**.
- Second variation is observed in 2021 Q3 was due to the release of a report by the National Security Commission on Artificial Intelligence (NSCAI). The **NSCAI report warned that AI could pose a serious threat to national security.**
- The other **variation during Q4 of 2022 was when ChatGPT was publicly released** which caused mostly positive and some negative sentiments at the time.

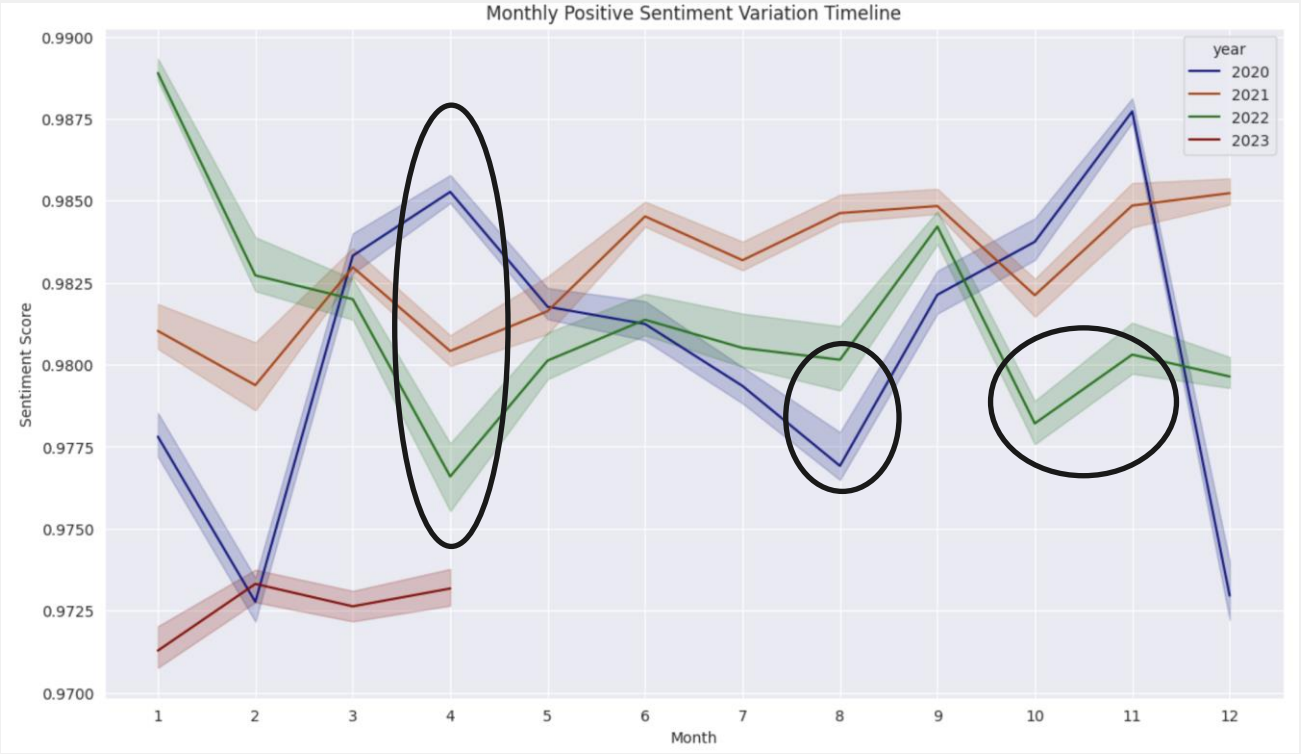


Sentiment Over Time – Analysis and Visualization

Positive Sentiments:

- The sentiments have been mostly positive during the 4 years timeline, while having variations when products got introduced.
- Examples: LAMDA (April 2022) got criticism for generating fake news or propaganda.
- (October 2022) Center for Responsible AI that found that AI is being used to track and surveil people around the world.

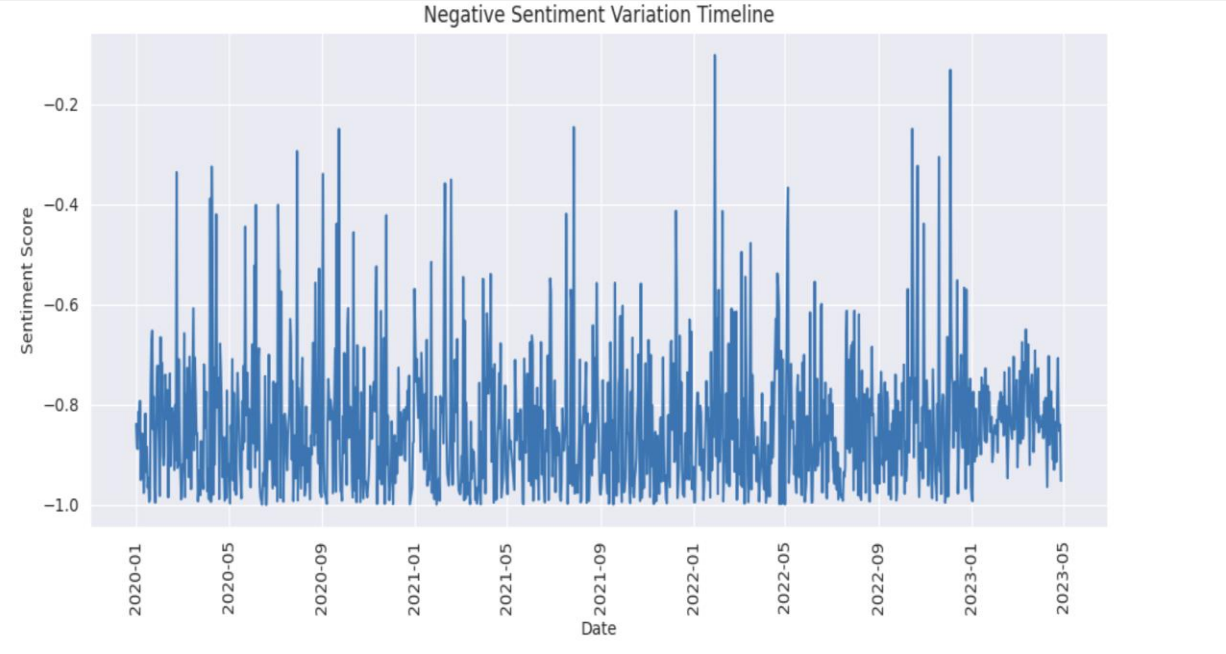


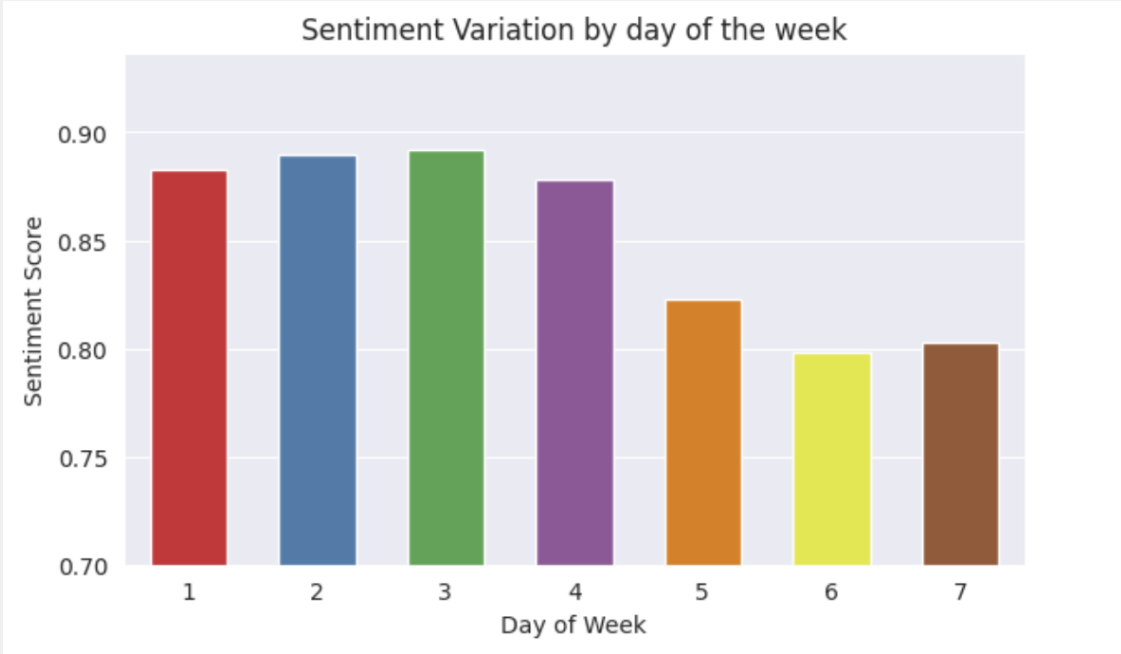


Sentiment Over Time – Analysis and Visualization

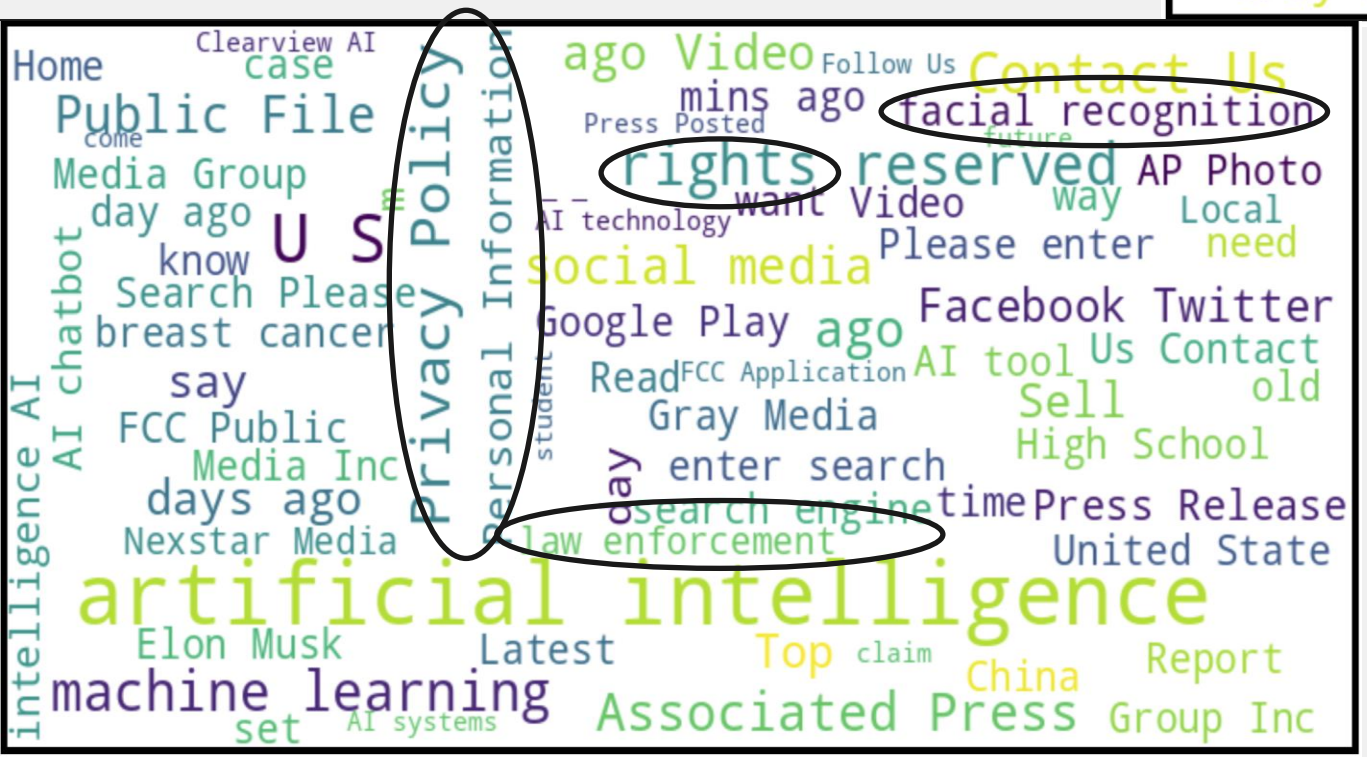
Negative Sentiments:

- The **negative sentiments in the first quarter of this year have been significantly reduced** with no major peaks. This is an indicator for the wide range of industries that started to evolve
- This can also indicate a **competition between various Generative AI** technologies like ChatGPT, Bard, etc.
- The **negative sentiments are observed mostly during** the later part of the weeks **(weekends)**.



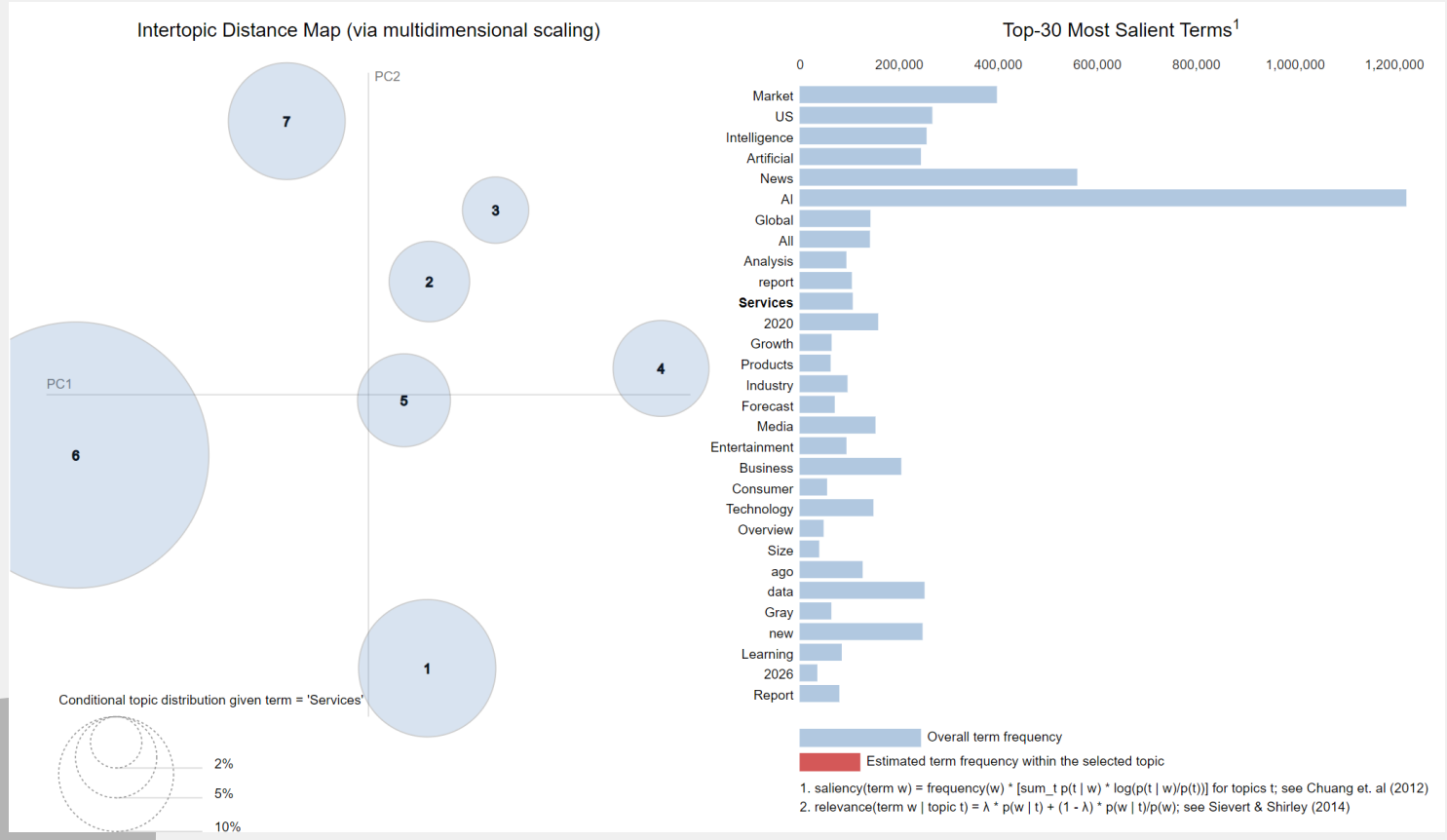


Positive Sentiment Word Cloud:
Most Common and Relevant Tokens



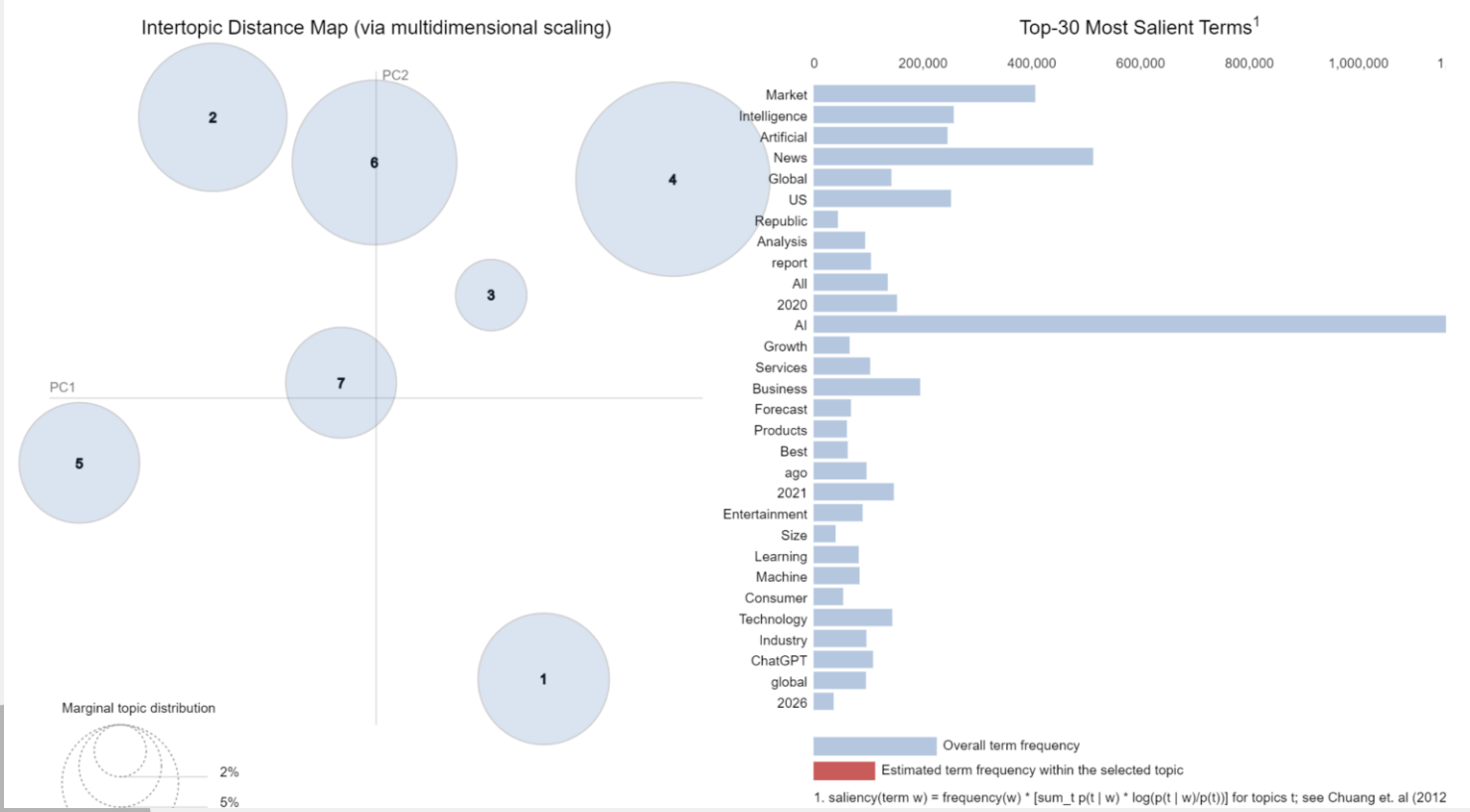
Negative Sentiment Word Cloud:
Most Common and Relevant Tokens

Topic Detection



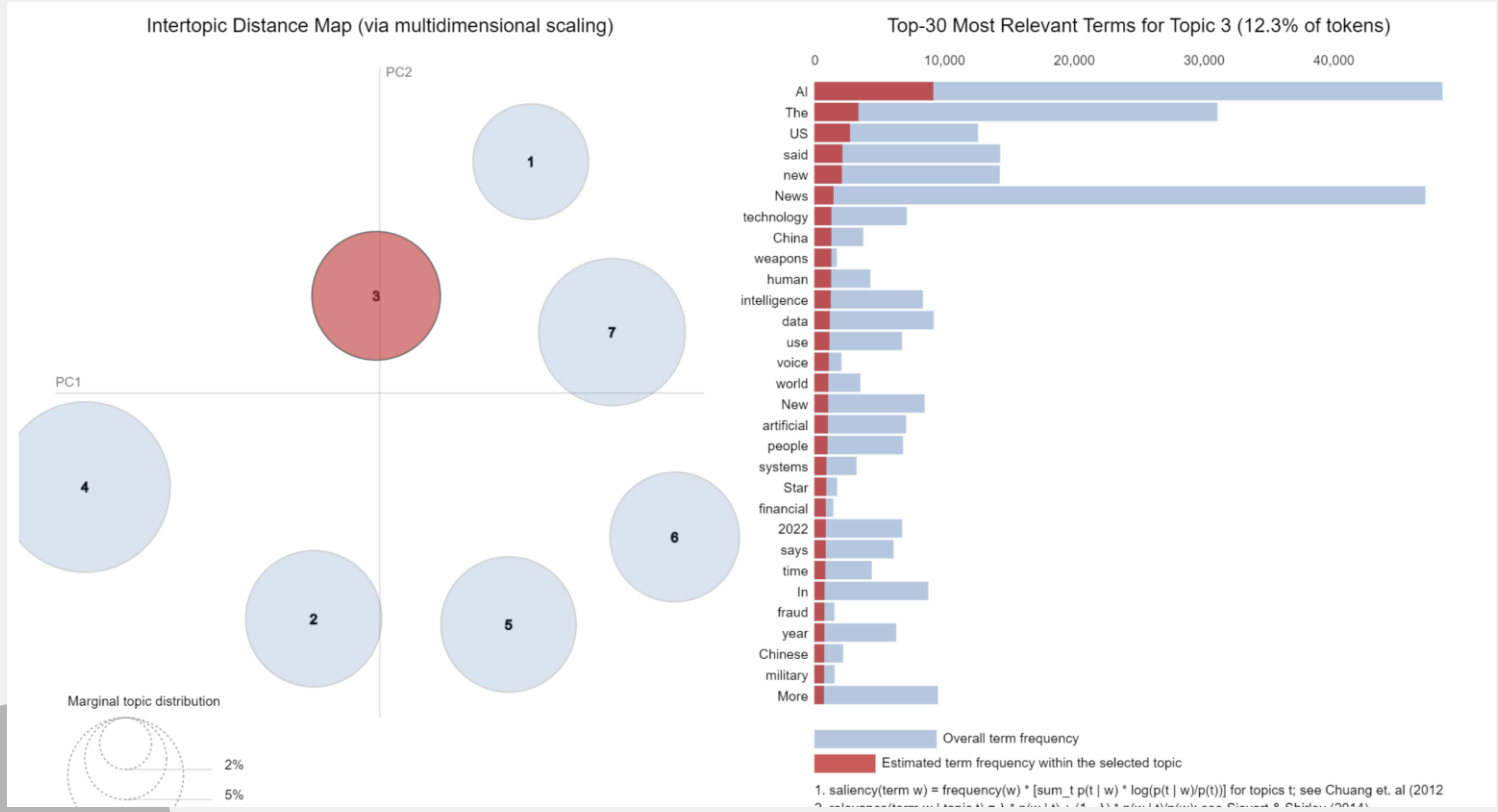
- Topic modelling was done using LDA to identify the topics in a collection of documents, and to assign each document to a topic.
- 7 Topic model captured topics with no overlap.
- **Using AI for capturing market trends**, growth projections, and competitive landscape.
- **AI Services, platforms, or tools** that businesses can leverage to automate processes.
- Significance of understanding how **AI technologies influence consumer** preferences.
- New technologies related to AI. It may include sectors like **autonomous vehicles, robotics, or AI-driven healthcare.**

Topic Detection : Positive Sentiments



- **Businesses can leverage market intelligence from AI services** to make informed decisions and identify potential investment areas.
- **Positive impact of AI on revenue generation**, operational efficiency, and overall business performance.
- **Highlighting new AI-driven tools**, algorithms, or platforms that have the potential to revolutionize industries and create new opportunities.
- Positive sentiment surrounding the capabilities and **potential applications of Chat GPT**.
- **Gray Television's new AI-powered ad platform** Gray IQ that uses AI to help advertisers target their ads more effectively.

Topic Detection : Negative Sentiments



- Highlights the negative sentiment surrounding the **potential for AI to be weaponized** or pose risks to security.
- Highlights the **negative reception of ChatGPT**.
- The need for careful consideration and **responsible implementation of new AI** technologies.
- **Potential usage of AI in fraudulent activities** and the importance of robust security measures.
- **AI's impact on healthcare** and the need for careful implementation to ensure patient safety and ethical considerations.

Entity Detection

- Entities from the news articles were detected using **SpaCy's NER** English pipeline.
- Top entities from Positive sentiment articles are:

ORG	ORG_Freq	LOC	LOC_Freq	People	People_Freq
Microsoft	78834	India	65850	Biden	7100
Google	66542	China	55739	Elon Musk	6168
OpenAI	45795	Us	50223	Musk	4836
Gray Television Inc	43343	UK	41680	Don	3785
IBM	29418	Japan	26384	Sundar Pichai	2936
GPT	26315	France	24888		
Amazon	22237	Canada	22995		
Bard	17583	Russia	22884		
Intel	13937	Germany	22494		
Apple	12978	Europe	21826		
NVIDIA	10656	United States	20840		

- Organization such **as Microsoft, Google, OpenAI, NVIDIA, Intel**, etc. which are mostly talked about are observed from the positive sentiment articles given their latest revolutionary research and products.

Example: **Gray Television popped here indicating their latest venture into advertising using AI**.

- **India was the most famous location** entity as it is a developing country and adapting to AI on a very fast pace.
- People like **Musk, Biden and Pichai** also come as the frequent "Person" entities

Entity Detection

- Top entities from Negative sentiment articles are:

ORG	ORG_Freq	LOC	LOC_Freq	People	People_Freq
Google	7826	US	15388	Jones	2800
Microsoft	7559	China	6029	Biden	1754
OpenAI	9657	India	3541	Musk	1530
Bing	1657	Russia	3335	Sam Altman	482
COVID	1617	UK	3120	Donald Trump	1312
Bard	1480	Florida	3050	Elon Musk	1179
CNN	1228	Ukraine	2673	Ai Weiwei	1125
Tesla	1104	New York	2556	Putin	879
Gray Television Inc	1065	California	2118	Rebekah Jones	879
Amazon	1050	Texas	1536	Joe Biden	742
Clearview AI	1914	United States	1417	DeStefano	738
The Associated Press	1960	Italy	1379		
Gray Media Group Inc	900	San Francisco	1181		
Reuters	860	Washington	1179		
Apple	811	London	1134		
Midjourney	702	Australia	1119		

- The organization **Google and Microsoft come again given their backlash on their AI** products and acquisitions.

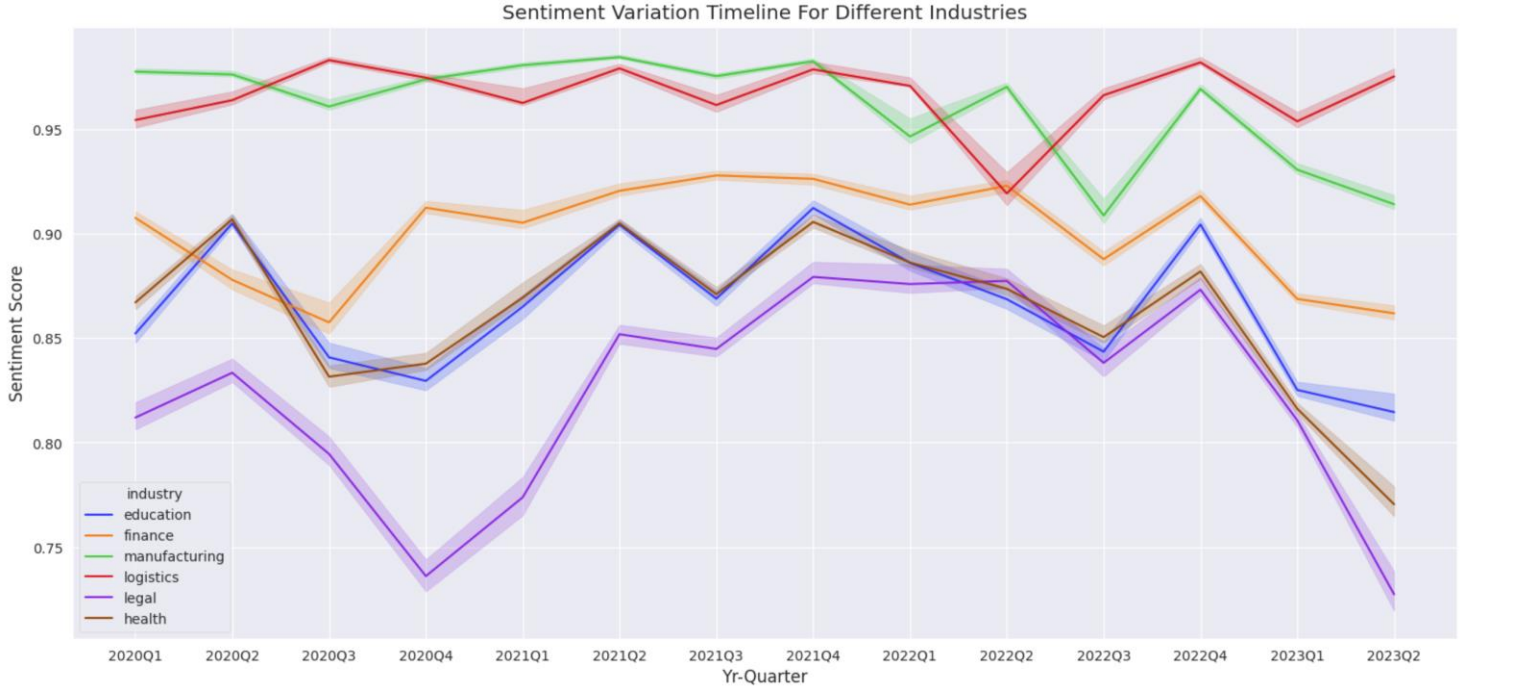
Example: **Bard got negative feedback when released and many people did not like Microsoft buying OpenAI.**

- Obviously, USA has been the center of negative sentiments given most development happened here.
- People like **Rebekah Jones, Sam Altman, Musk, Trump,** etc come up here.

Example: **Rebekah Jones's controversies involving misleading COVID-19 data reports**.

Targeted Sentiment Identification

- Industry level Sentiment Variation is identified to check which industries have more effects:



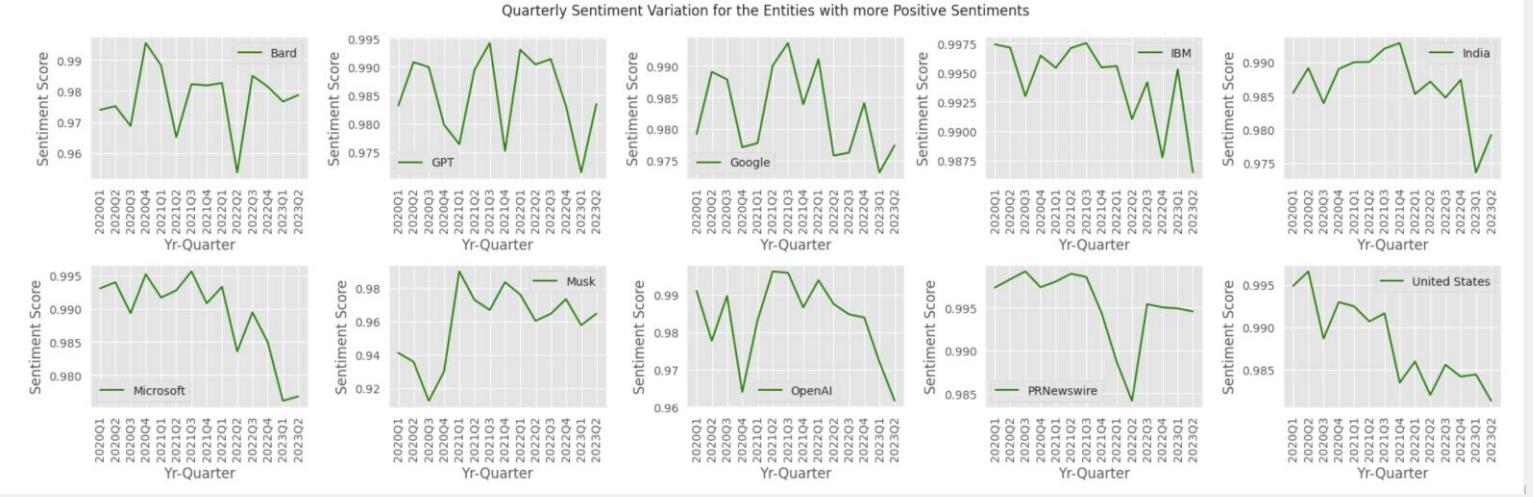
- From the plot it can be seen that the **logistics and manufacturing industries are the top industries** which are trying to carry the positive sentiment and automate their processes for **increased productivity and revenues**.
- There is **not a good reception of AI in both education and healthcare** given the issues with data governance, quality and privacy.
- AI has seen a **rise in being used by legal professionals**, most likely NLP applications for doing the work of paralegals.
- **Finance industry has been consistently adapting to the trends in AI** and is carrying the positive sentiment with not many ups and downs.

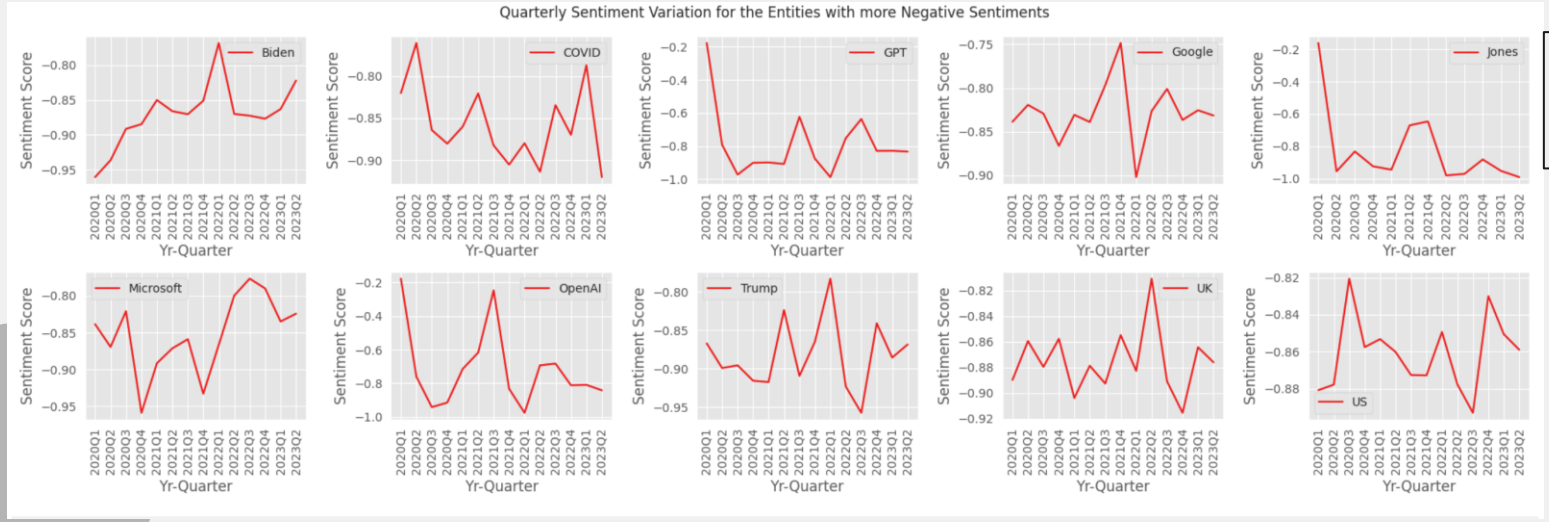
Targeted Sentiment Identification

Entity level sentiment variation timeline for Positive Sentiment Articles

- Most of the “ORG” entities here have not been consistent with maintaining sentiments over the past 4 years.

Example: Sentiment for ChatGPT has been going up and down continuously for every update it releases.





Entity level sentiment variation timeline for Negative Sentiment Articles

- Negative sentiments for Microsoft and Google are following a somewhat similar pattern, indicating their competition in developing newer AI technologies and the setbacks.
- US and UK seen negative sentiments much higher than other “GPE” entities.

Thank You!