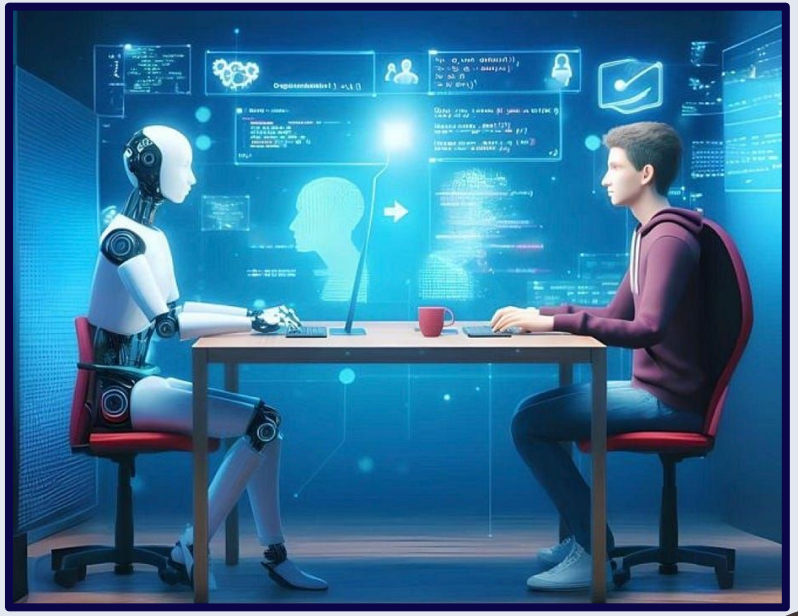
AI Impact Analysis: Industries & Jobs



Tsengee Sundui
Dec 7, 2023

Table of Contents

Executive Summary

Project Overview:

This project investigates the recent impact of artificial intelligence (AI) on industries and jobs, utilizing a vast dataset of ~200,000 news articles focused on Data Science, Machine Learning, and Artificial Intelligence.

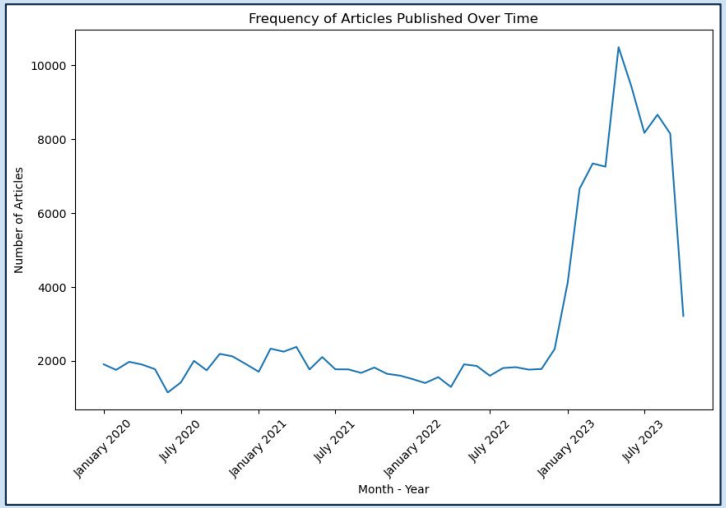
Through extensive data preprocessing and advanced natural language processing techniques, the aim is to extract meaningful insights and provide actionable recommendations for navigating the AI revolution.

Key Findings:

- Overall Positive Sentiment
  - The general sentiment in articles related to AI, Data Science and ML is positive
- Industry-Specific Positivity
  - Finance and technology sectors exhibit higher positivity towards AI
- Challenges and Negativity
  - Negative sentiment likely stems from ethical concerns, privacy issues, and potential misuse of AI
- ChatGPT's Impact
  - Introduction of ChatGPT initially led to a sentiment decrease
  - This decline was temporary, attributed to initial fears, and did not persist
- Diverse Sentiments exist within the same industry (i.e. Healthcare AI)
  - Positive sentiments in healthcare AI relate to innovations in diagnostics, personalized treatment, and improved patient care
  - Negative sentiments may arise from concerns about data privacy, potential misuse of AI algorithms, etc.

Data Profile

- 199,677 news articles (size ~900 MB)
- Key features: '*date*', '*title*', '*text*'
  - converted date to the correct data type
- Other features (dropped): 'url', 'language'
  - all the articles were in English



Cleaning & Preprocessing Steps

- Dropped rows with duplicate titles
  - data reduced to 138,841 articles
- Removed unnecessary phrases from text
  - such as 'Skip to content', 'Sign up', or 'Leave a Reply', etc.
- Removed special characters, extra spaces, URLs, email addresses, non-printable characters, and non-ASCII characters
- Removed stopwords, single-character tokens, and punctuation
- Applied Lemmatization and Tokenization
- Filtered by most relevant topics (LDA)
  - dataset reduced to 110,065 articles
- Kept the articles shorter than or equal to the 80th percentile length
  - data further reduced to 88,056 rows

10% of the data was randomly sampled for faster processing. Perplexity and coherence scores were used to determine the optimal number of topics for the news articles, using Latent Dirichlet Allocation (LDA)

	Potential Theme	Keywords
Topic 0	General news topics	'news', 'hour', 'video', 'weather', 'local', 'day', 'stories', 'sports'
Topic 1	Artificial intelligence and technology	'ai', 'new', 'chatgpt', 'google', 'tech'
Topic 2	Entertainment-related	'star', 'show', 'year'
Topic 3	Best buys, insurance, cryptocurrency	'best', 'buy', 'insurance', 'crypto', 'online', 'ai', 'bitcoin'
Topic 4	General topics related to US media	'us', 'news', 'media', 'services', 'technology', 'business', 'policy'
Topic 5	Technology and business-related	'ai', 'data', 'technology', 'business', 'learning', 'machine', 'customer'
Topic 6	Market reports and analysis	'market', 'report', 'global', 'analysis', 'growth', 'research', 'ai'
Topic 7	News and community-related	'news', 'public', 'radio', 'music', 'maine', 'schedule', 'community', 'ai'
Topic 8	General news and health	'news', 'republic', 'email', 'ai', 'world', 'privacy', 'technology', 'health'
Topic 9	Stock markets and prices	'stock', 'market', 'ai', 'share', 'company', 'price', 'stocks', 'nasdaq'

Filtered the data by the most relevant topics

To help identify what industries and jobs would be most impacted by AI, the following topics seemed most relevant to include:

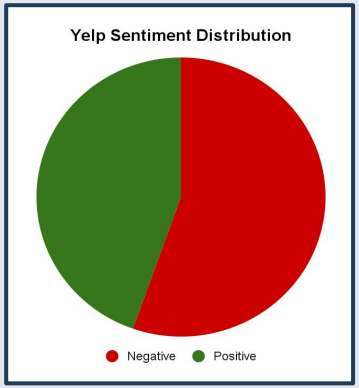
- Topic 1: AI and tech
- Topic 3: Best buys and cryptocurrency
- Topic 5: Broad technology and business-related topics
- Topic 6: Market reports and analysis
- Topic 8: News and technology-related topics, including privacy and health
- Topic 9: Stock markets

After discarding the less relevant topics, the dataset decreased from 138,841 news articles to 110,065.

Sentiment Analysis: Yelp and VADER

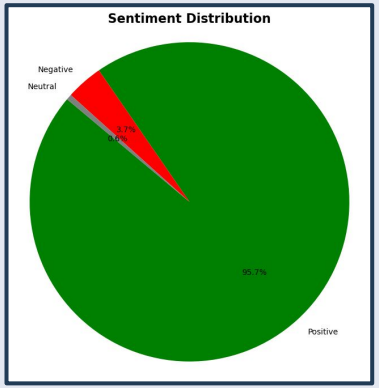
1st approach: **Yelp** Sentiments

- Custom sentiment prediction model using Yelp open-source data
- Result: a positive or negative sentiment category for text (binary classification)
- The Yelp model classified more than half of the articles as negative, signaling the need for an additional approach

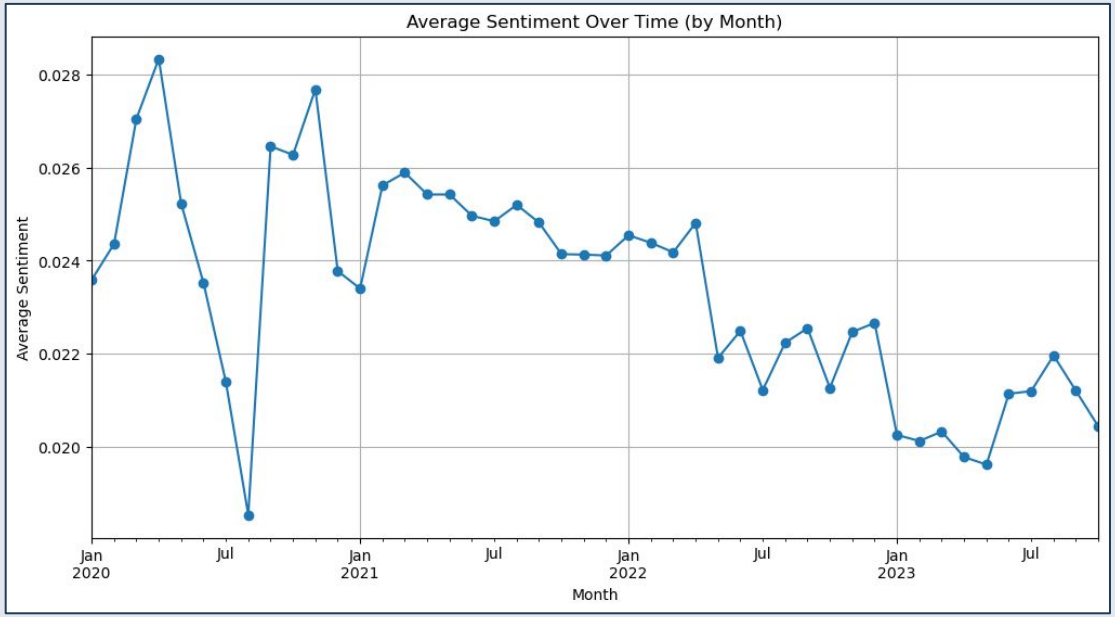


2nd approach: **VADER** (Valence Aware Dictionary for Sentiment Reasoning)

- Result: each word is assigned a numerical score indicating its polarity (positive or negative) and intensity
- Calculated average sentiment scores, and used it to categorize the news text into positive, negative (or neutral) sentiments



How the general sentiment changed over time:

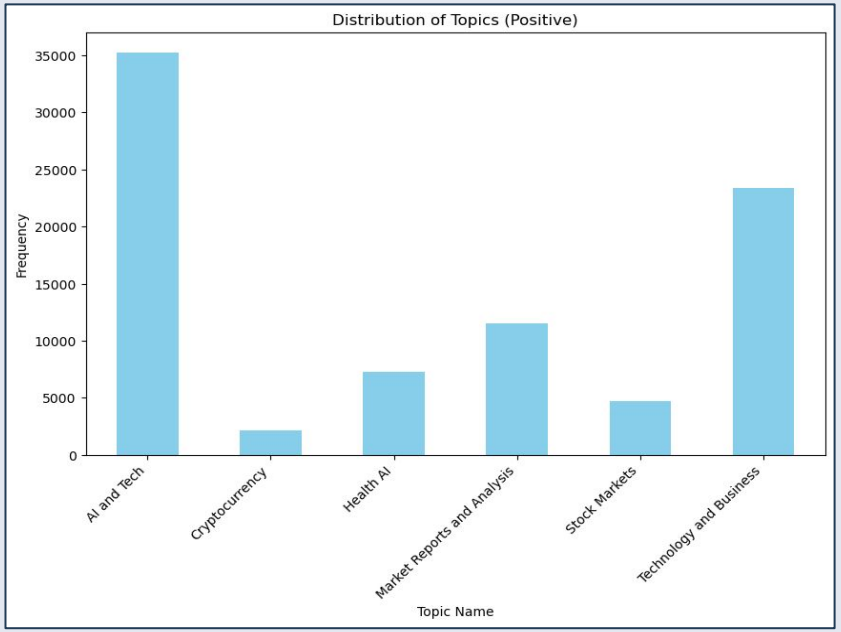


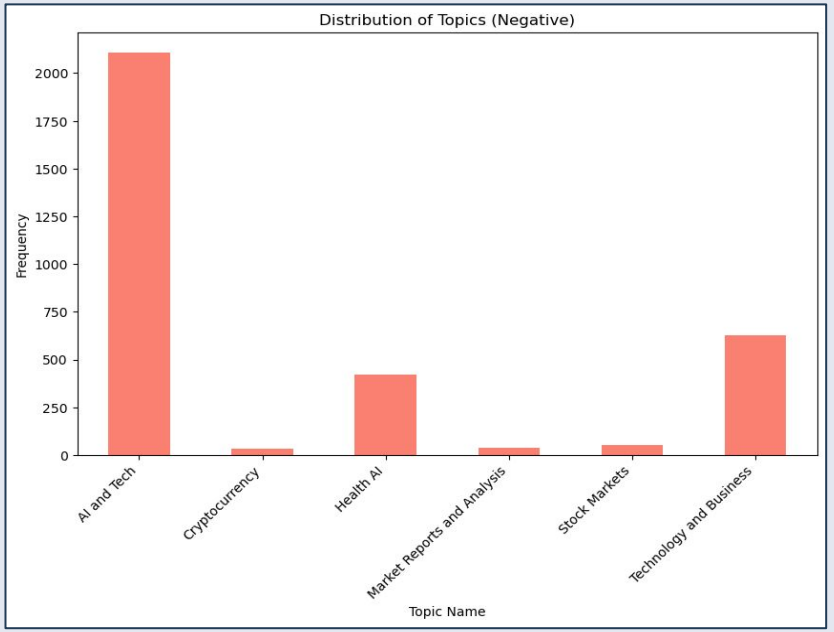
Using the VADER approach, it can be seen that the **overall sentiments were positive** for the news articles (above 0).

But, there were some **fluctuations**:

- Sharp decline in April, 2020
  - COVID-19 initial peak
- Increased during Aug-Nov, 2020
  - AI driven developments in healthcare and vaccination
- Further decline in April, 2022 and December, 2022
  - Hiring freezes started across big tech companies (data science related jobs) in April
  - ChatGPT was released in Nov, 2022 (there may have been some initial fears and hesitancy)

Topic Frequency for Positive vs Negative Sentiments



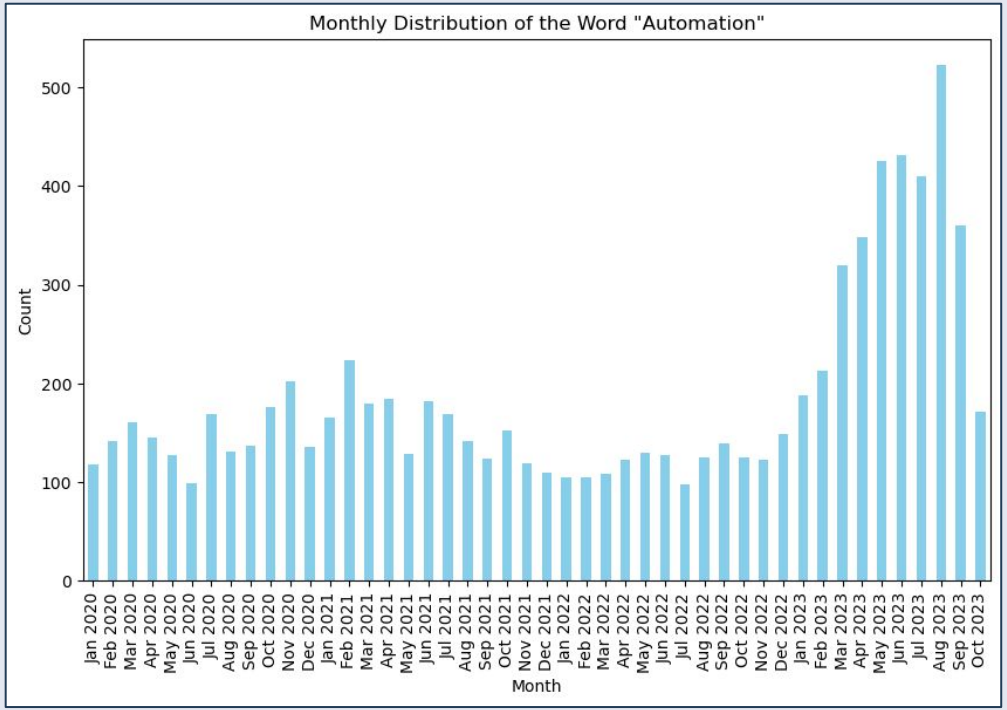


Topics such as AI and ChatGPT (Tech) generated both positive and negative sentiments. Whereas, news about cryptocurrency, financial market analysis and stock markets had more positive sentiments overall. Both positive and negative sentiments touch on health-related topics.

Sentiments: Key Insights

- Diversity of Topics
  - Both positive and negative sentiments cover a diverse range of topics, including finance, technology, health, and global events
- Financial Market Focus
  - Positive sentiments include topics related to the financial market, stock prices, and global market analysis, suggesting positive reactions to financial news
- Health and Technology Challenges
  - Positive sentiments focus on healthcare
  - Negative sentiments address challenges in health and machine learning
- Technology Innovation
  - Positive sentiments express excitement about new technologies, AI, and innovation
  - Negative sentiments include concerns about technology, companies, and market-related challenges.

The Impact of AI Advancements: ChatGPT



In March of 2023, Goldman Sachs published a report, indicating that ~25% of the tasks in US and Europe can be automated using AI.

After the launch of ChatGPT in **November of 2023**, we can see that the frequency of word 'automation' in news articles skyrocketed.

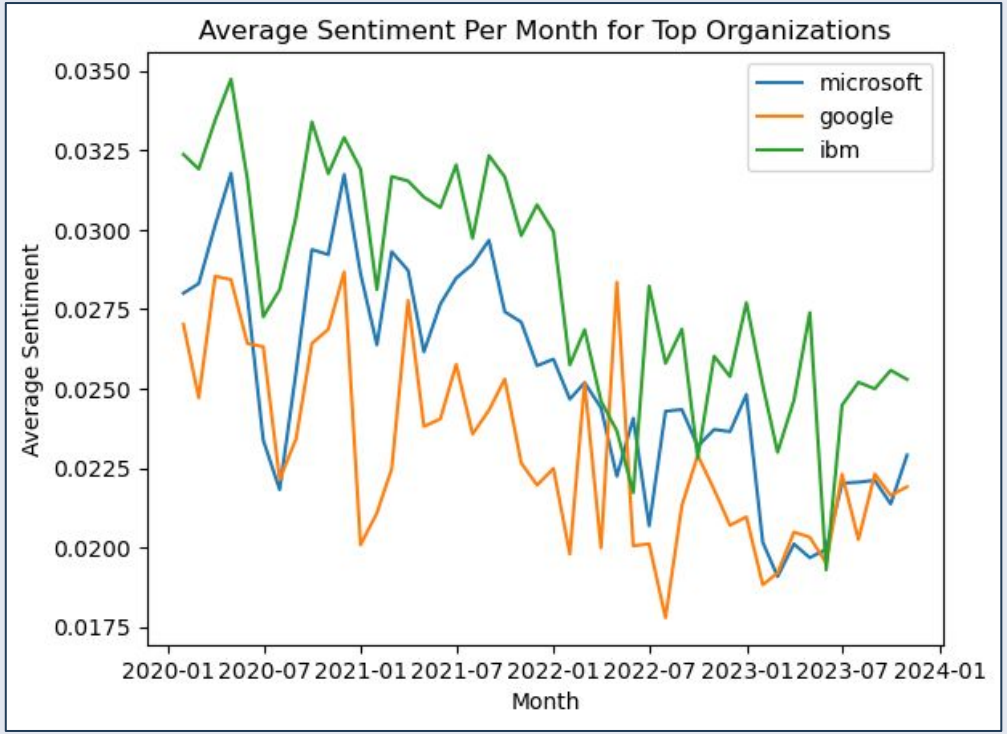
Positive sentiment over time for top organizations

Microsoft, Google and IBM were the top three most mentioned entities with positive sentiments in the news articles.

Microsoft is a major investor and partner for Open AI, who launched ChatGPT.

Google's parent company, Alphabet, also launched a conversational AI chatbot named Bard (as a response to ChatGPT). And, Google DeepMind focuses on AI research.

In May 2022, IBM announced that the company had signed a multi-year Strategic Collaboration Agreement with Amazon Web Services to make a wide variety of IBM software available as a service on AWS Marketplace.



Different sides of the AI landscape: Healthcare

Positive Sentiment Example

"COVID-19 Update: Global AI In Medical Imaging Market is Expected to Grow at a Healthy CAGR..."

- Date: 2020-12-03
- The article reports that the global AI in Medical Imaging market is growing, citing major players such as General Electric, IBM Watson Health, Philips Healthcare, Samsung, etc..

Negative Sentiment Example

"Will AI really destroy humanity? - Digital Journal"

- Date: 2023-06-27
- The article discusses potential risks and uncertainties associated with the evolving field of artificial intelligence, and the misuse of AI by "bad actors" and risks of AI machines with self-preservation goals

These two articles are a good example of the presence of positive and negative discussions surrounding AI within the same industry.

AI's Impact on Industries and Jobs: Discussion

Financial Market Focus:

- Finance and insurance industries are likely embracing AI for market analysis and decision-making processes.
- This trend may result in increased demand for data scientists and analysts, while roles involving routine tasks like data entry may diminish.

Health and Technology Challenges:

- There may be some hesitancy in the healthcare industry to adopting AI due to ethical risks.
- However, the introduction of AI in health may create roles such as health data analyst or AI implementation specialist.

Technology Innovation:

- Industries expressing excitement about new technologies and innovation, such as technology and AI companies, are likely to create roles in research and development.
- This could include positions like AI researcher or technology innovation manager.

Navigating the AI Revolution

01	Invest in AI Education	• Encourage continuous learning and upskilling among your workforce to understand AI technologies
02	Human-Machine Collaboration	• Emphasize the role of AI as a tool to augment human capabilities rather than replace jobs
03	Monitor Industry Trends	• Stay up to date on the AI trends and advancements within your industry
04	Create roles to address challenges	• Roles in technology risk assessment or market strategy analysis might see a rise

Conclusion

In summary, the prevailing sentiment toward AI is predominantly positive, particularly in industries such as finance and technology where the adoption of AI is more widespread.

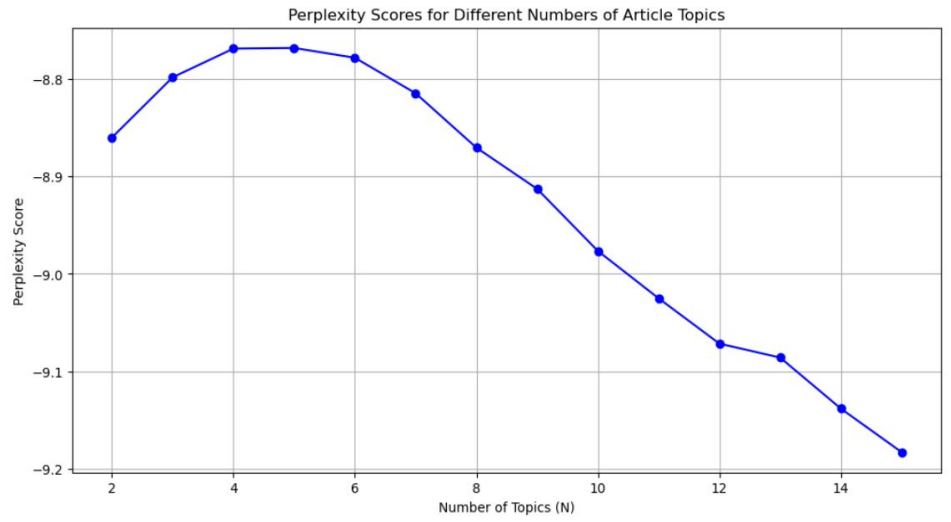
However, negative sentiment arises from ethical concerns, privacy issues, and the potential misuse of AI technologies.

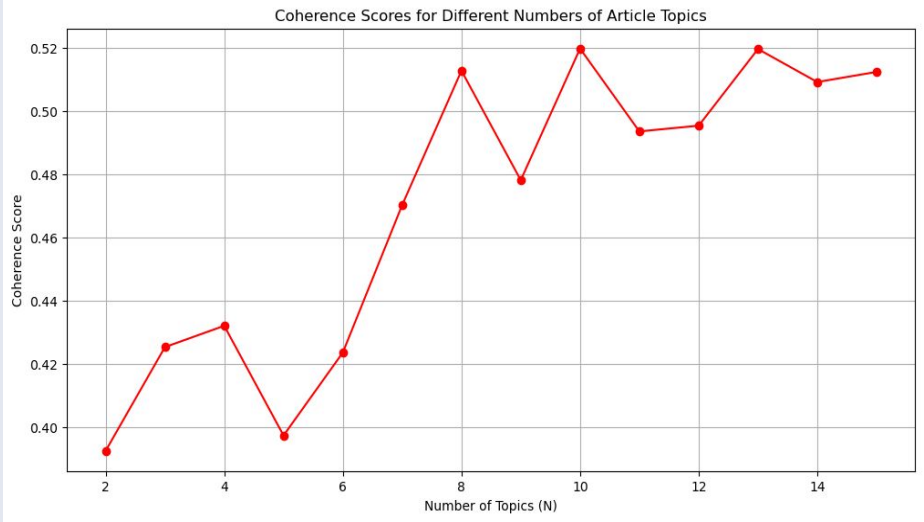
Notably, the introduction of ChatGPT initially led to a decrease in sentiments, which could be attributed to fears and hesitancy surrounding its capabilities, including concerns about improper use in educational settings, apprehension regarding data collection and privacy issues

- It's important to highlight that this decline in sentiment was temporary and did not persist, indicating a potential shift in perceptions over time.

Appendix

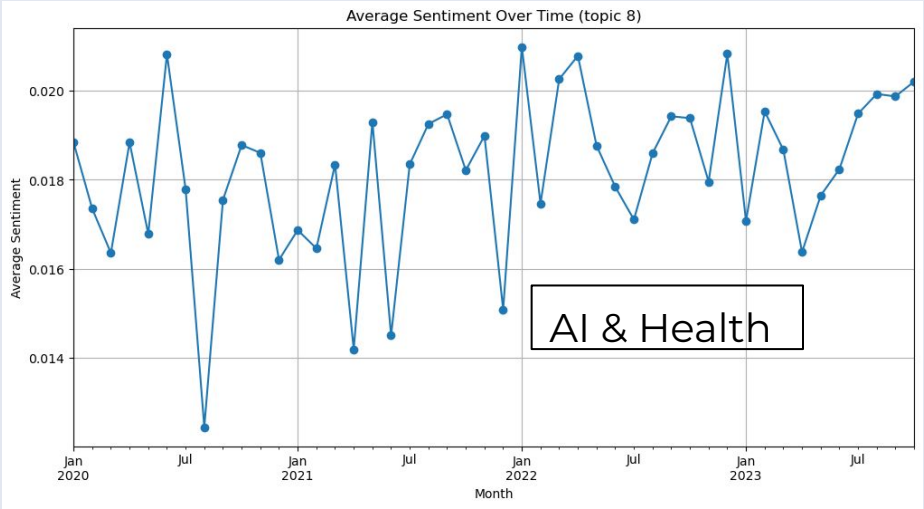
Perplexity and Coherence Scores

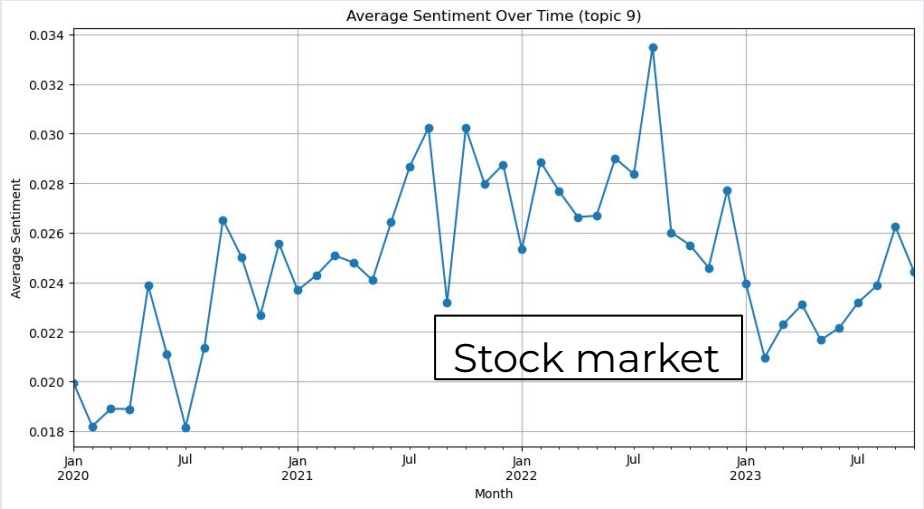


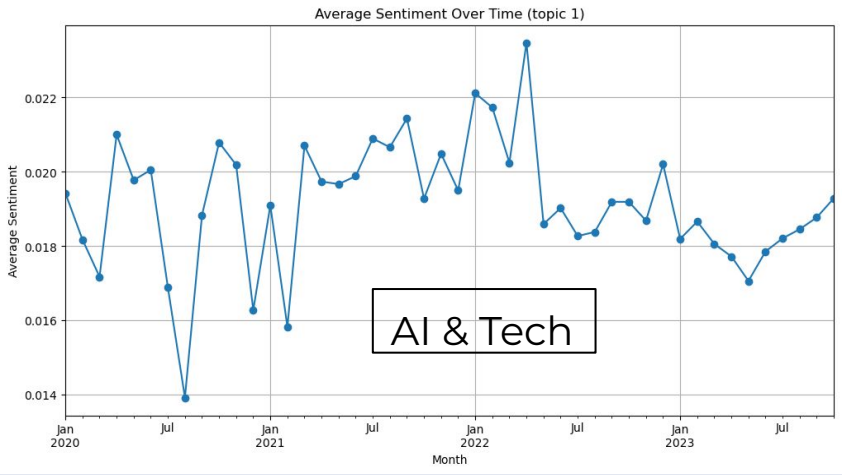


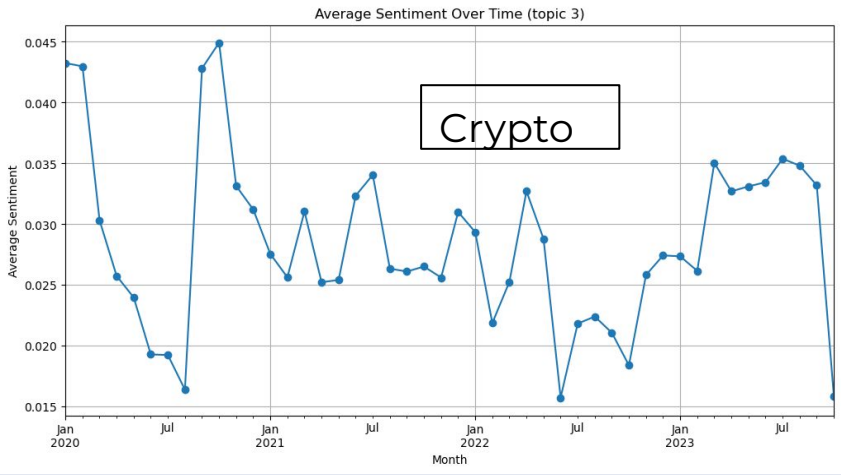
A lower perplexity score (less negative value) means the topics are more informative. A higher coherence score suggests a better quality of topics. For the news articles, the optimal N seems to be at **10 topics** (coherence peaks and perplexity is lower). 13 topics might be suitable as well, but I chose to go with 10 instead, since the coherence scores are the same and perplexity only goes down a little more.

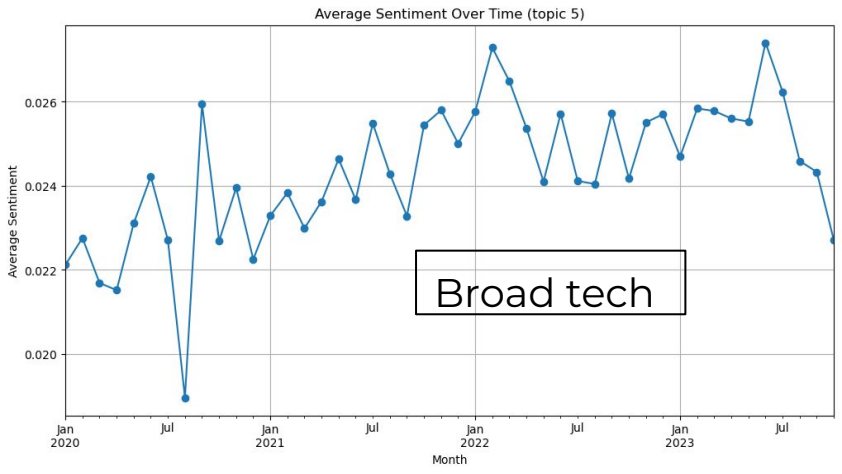
Average Sentiments by Topic

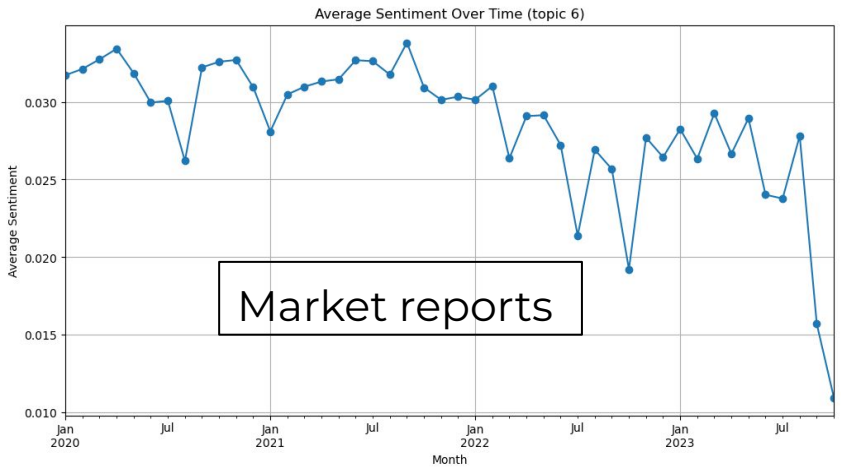


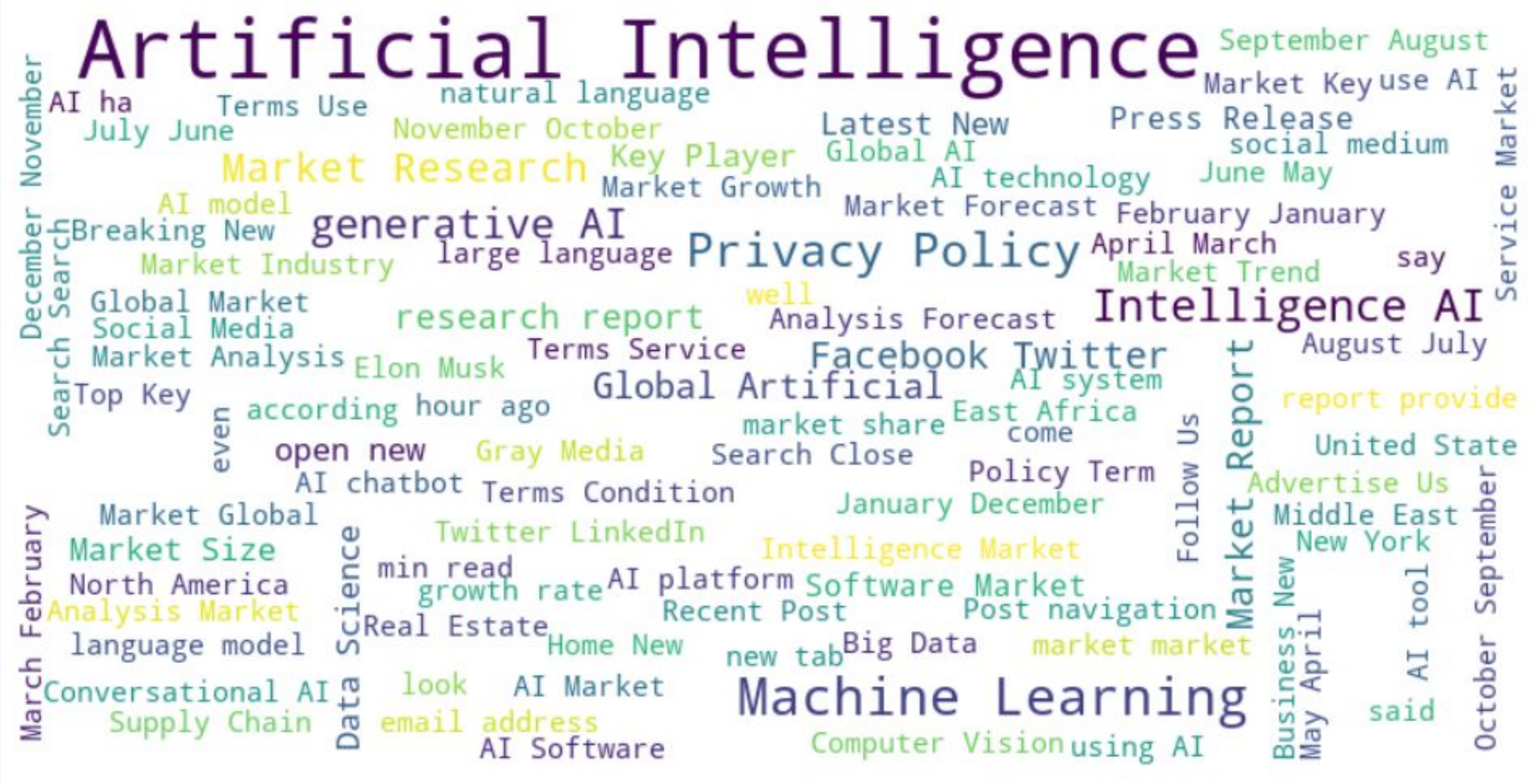



Average Sentiment Over Time (topic 1)
Average Sentiment
Month
AI & Tech


Average Sentiment Over Time (topic 3)
Average Sentiment
Month
Crypto


Average Sentiment Over Time (topic 5)
Average Sentiment
Month
Broad tech


Average Sentiment Over Time (topic 6)
Average Sentiment
Month
Market reports


Artificial Intelligence
September August
Market Key
use AI
November
AI ha
Terms Use
natural language
July June
November October
Latest New
Press Release
social medium
Market
Market Research
Key Player
Global AI
AI technology
June May
AI model
Market Growth
Market Forecast
February January
December
Search
Breaking New
generative AI
large language
Privacy Policy
April March
Market Industry
Market Trend
say
Service
well
Global Market
Social Media
research report
Analysis Forecast
Intelligence AI
Search
Market Analysis
Elon Musk
Terms Service
Facebook Twitter
August July
Top Key
Global Artificial
AI system
report provide
even
according
hour ago
market share
East Africa
come
Market Report
open new
Gray Media
Search Close
Follow Us
United State
AI chatbot
Terms Condition
Policy Term
Advertise Us
March February
Market Global
January December
Middle East
Twitter LinkedIn
Market Size
Intelligence Market
New York
Data Science
min read
North America
AI platform
Software Market
growth rate
Business New
AI tool
October September
Analysis Market
Real Estate
Recent Post
Post navigation
language model
Home New
Big Data
market market
new tab
May April
Conversational AI
look
AI Market
Machine Learning
said
Supply Chain
email address
Computer Vision
using AI
AI Software