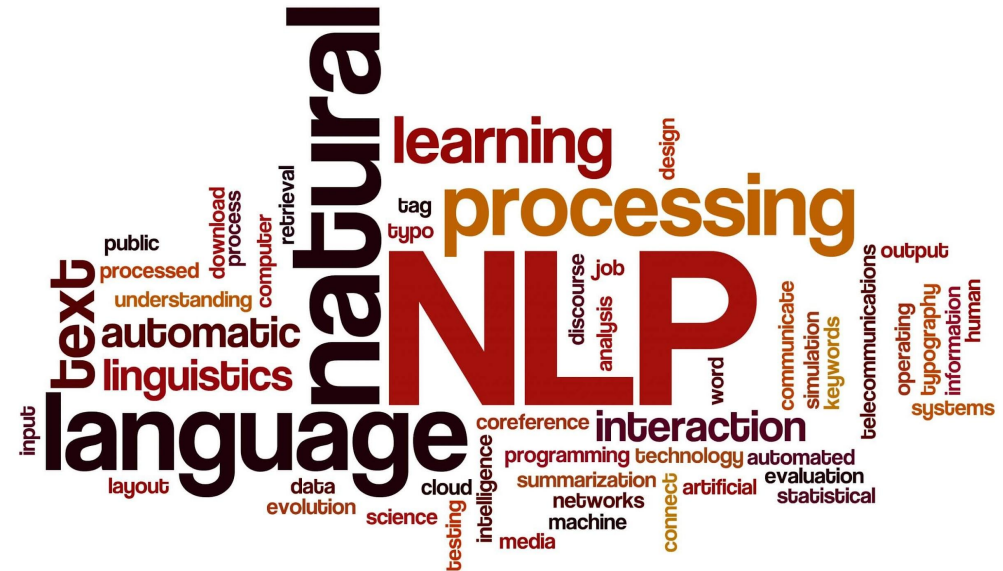# Unveiling the Influence of Artificial Intelligence in Industries: Harnessing Natural Language Processing for Impact Analysis

Code Reference: Github Link

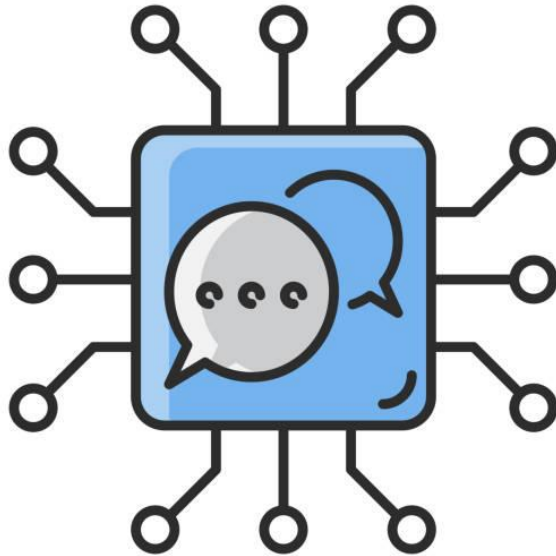**Natural Language Processing, Spring 2023**

Ankita Nambiar

ankita7@uchicago.edu

THE UNIVERSITY OF CHICAGO

Agenda

Executive Summary
Data Profile
Preprocessing
Modelling Workflow
Important Topics
Sentiment Analysis
Conclusions and Recommendations

Executive Summary

The objective of this project is to identify the industries and job lines that are most likely to be impacted by Artificial Intelligence (AI) and Machine Learning in the next several years and provide actionable recommendations on how to automate tasks and improve employee productivity.

The analysis was conducted using the text of approximately 200,000 news articles on topics related to Data Science, Machine Learning, and Artificial Intelligence. Natural Language Processing (NLP) techniques were applied to extract meaningful insights from the unstructured text.

Background

Artificial Intelligence (AI) has become a transformative technology with numerous applications, including medical diagnosis, fraud detection, and chatbots.

AI has made significant advancements due to the development of large language models (LLMs). The release of ChatGPT has contributed to a boost in the development and adoption of AI technologies.

These advancements opened up new possibilities for automation, improved productivity, and the creation of intelligent systems.

Findings

Industries that rely on human touch, such as education, news, media, and art, express concerns about the potential loss of human sentiment, emotion, creativity, and judgment. Conversely, industries with prior experience in handling data, like business, finance, healthcare, and insurance, exhibit more positive sentiments towards AI.

Different industries have varying sentiments towards AI based on their job functions and experiences. In industries accepting AI, there will be a reduction in mundane jobs and manual labor.
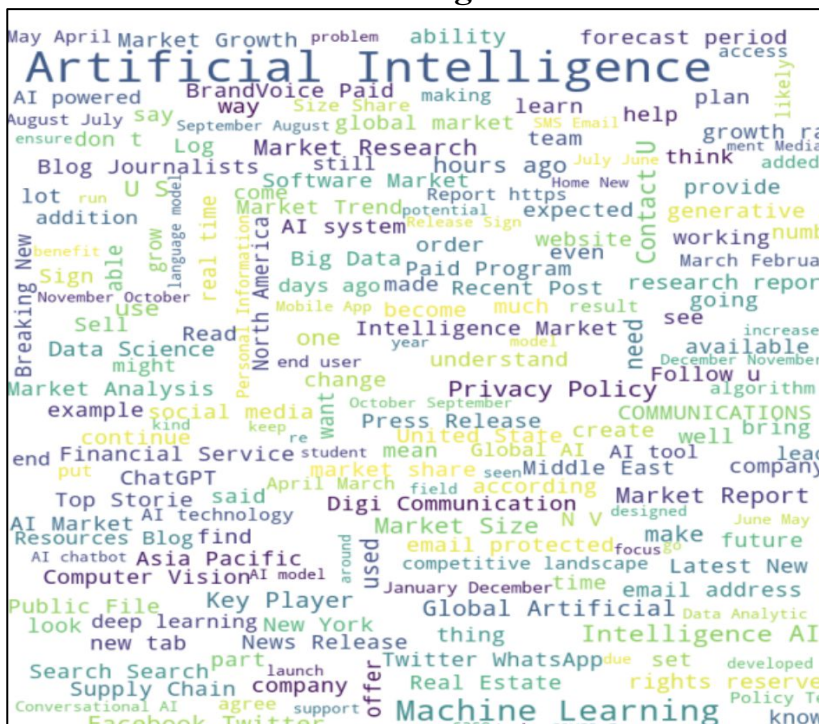
Data Profile

Description:
The dataset consists of 200,332 news articles on topics related to Data Science, Machine Learning, and Artificial Intelligence. The dataset is a size of ~ 900 MB.
Source:
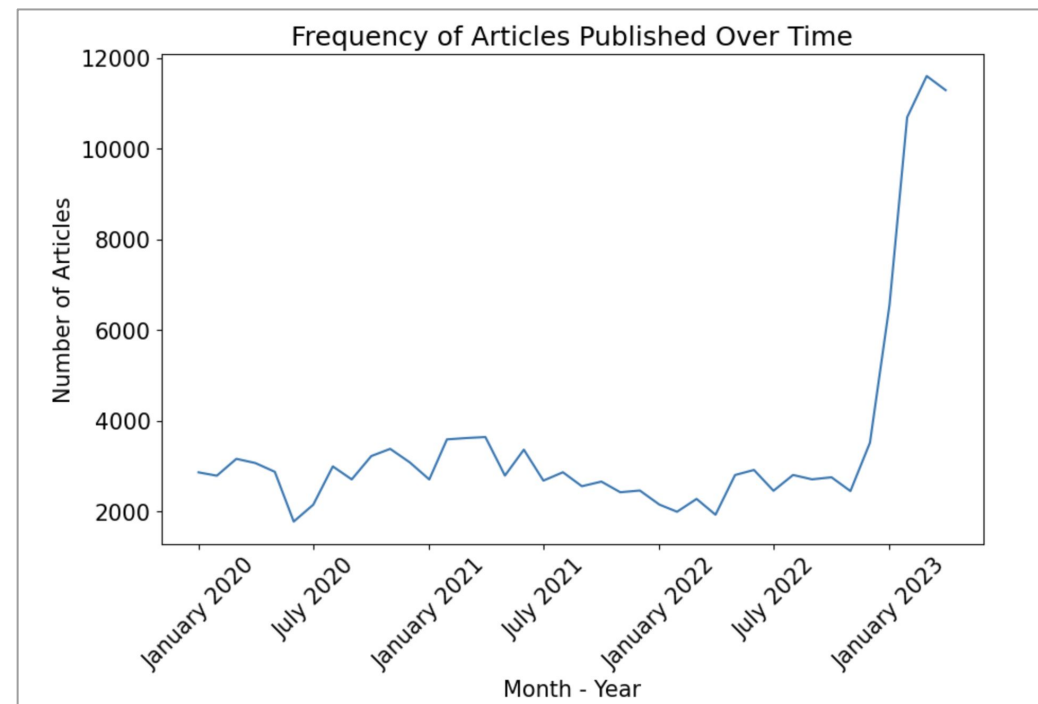Professor Nick collected articles for the dataset.

Features:
- url: link to article
- date: publication date, from '01-01-2020' to '04-28-2023'
- language: primary language of article
- title: unstructured title of article
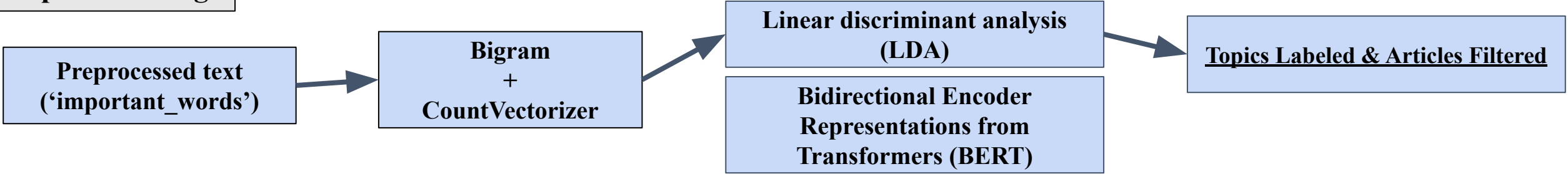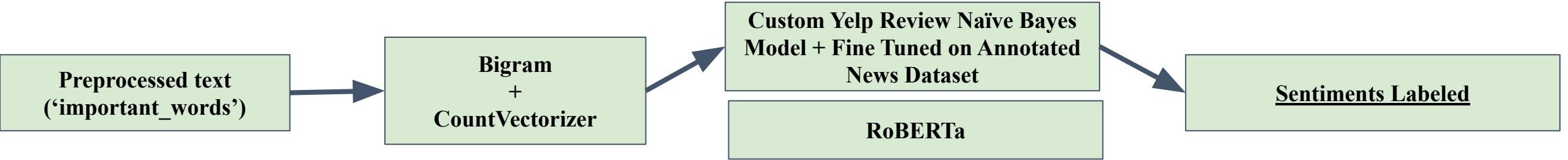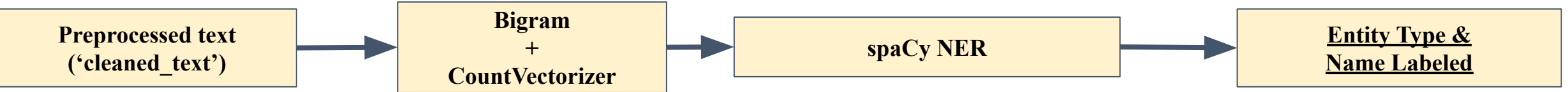- text: unstructured text of article

Word Cloud on Original Text

Methodology: Preprocessing

Disregard Unnecessary
- Non-English articles → 200,332 articles in sample
- Duplicates → 140,223 articles in sample
- Feature 'title': Feature 'text' has the title within

Clean-up Noise
- Eliminate newlines, tabs, extra spaces, remnants of web crawls, and username mentions.
- Delete words after common article ending phrases (if placement is at the end of the article).
  - Ex: 'Related Articles:', 'Copyright', 'For additional information'
- After, remove words less than 25 characters and sentences less than 5 or greater than 45 words → most likely missed spam or messy text.

Create Features
- 'important_words'
  - stop words removed
  - special characters removed
  - lowercase
  - lemmatized
- 'cleaned_text': Used when context is needed (Ex: Entity Recognition)

Word Cloud On Preprocessed Text

Effects of Preprocessing

	Average Sentence Length	Average Article Length
Original	61.3 Words	1330 Words
Cleaned	31.7 Words	540 Words

Methodology: Modelling Workflow

Topic Modeling

Preprocessed text ('important_words') → Bigram + CountVectorizer → Linear discriminant analysis (LDA) / Bidirectional Encoder Representations from Transformers (BERT) → **Topics Labeled & Articles Filtered**

Sentiment Analysis

Preprocessed text ('important_words') → Bigram + CountVectorizer → Custom Yelp Review Naïve Bayes Model + Fine Tuned on Annotated News Dataset / RoBERTa → **Sentiments Labeled**

Named Entity Recognition

Preprocessed text ('cleaned_text') → Bigram + CountVectorizer → spaCy NER → **Entity Type & Name Labeled**

Important Topics in the Articles: AI in Business & LLM applications

Topic Frequency in News Articles
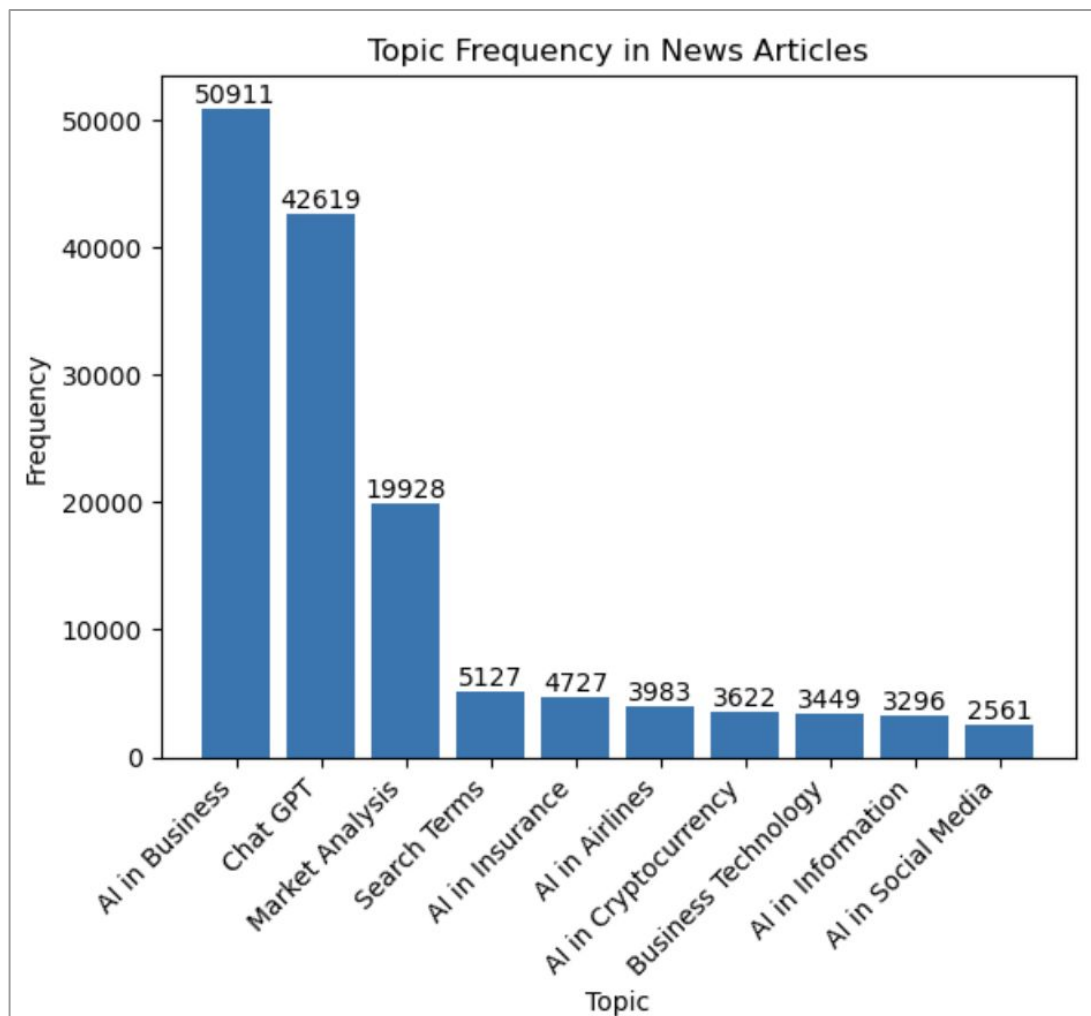
Filtered out Articles in Topic 3 'Search Terms,' as it is not AI / data science related.

0. AI in Business
❏ Emerging trends in AI adoption by businesses across industries.

1. ChatGPT
❏ Exploring the capabilities and limitations of AI language models like ChatGPT.

2. Market Analysis and Research
❏ Leveraging data analysis and market research for informed decision-making.

4. AI in the Insurance
❏ AI applications in automating insurance processes and risk assessment/ fraud detection in the insurance sector.

5. AI in Airlines
❏ Opportunities and challenges in implementing AI solutions in the airline sector.

6. AI in Cryptocurrency
❏ Using algorithms for Trading from Investing to Fraud Detection.

7. Business Technology
❏ The integration of AI into various aspects of business technology like recommendation systems.

8. AI in Information
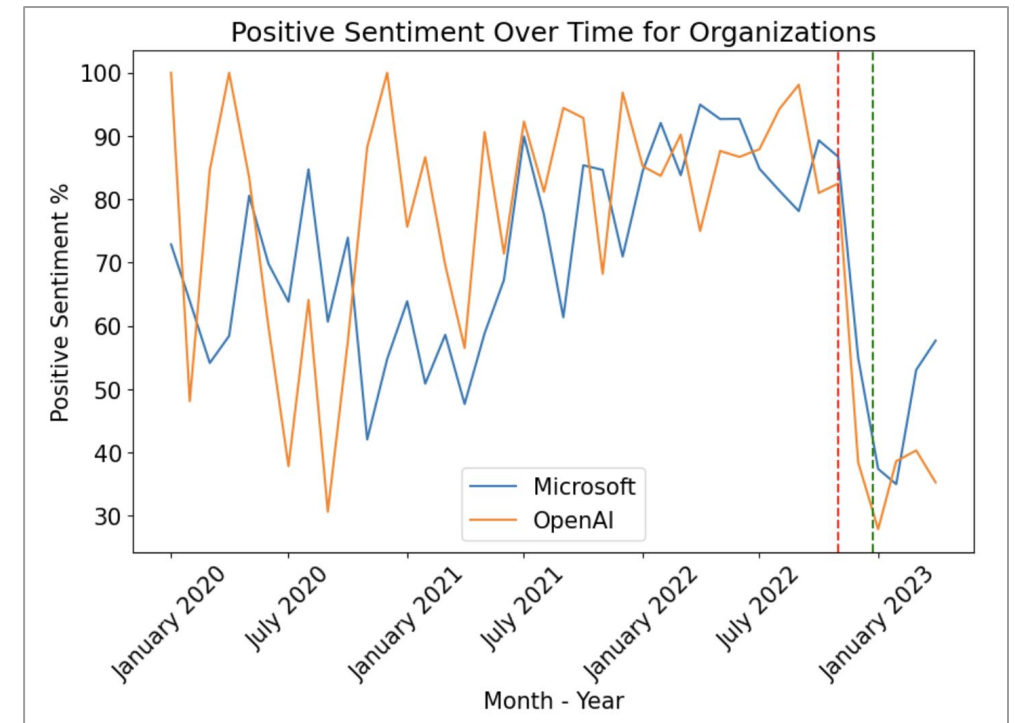❏ The application of AI in the field of information management in automating decision-making techniques.

9. AI in Social Media
❏ Using AI for facilitating personalized user experiences, content recommendation, and targeted advertising.

Change in AI Sentiment due to ChatGPT
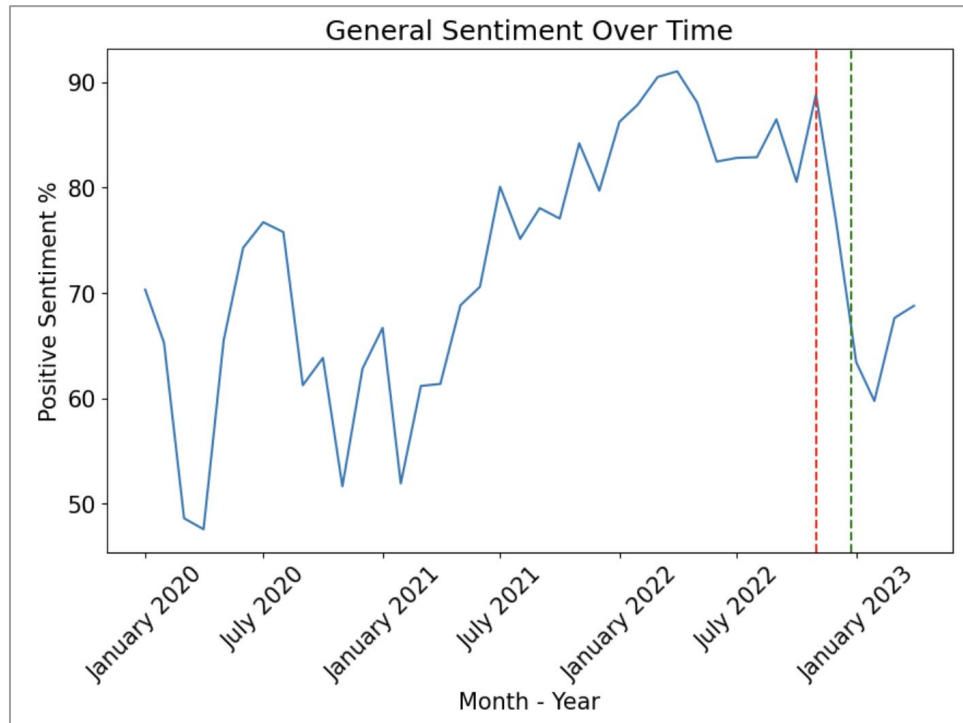
Sentiment % = 100 * # of Positive Article in the Month / # of Total Articles in the Month

After the **release of ChatGPT** in November 2022, the overall sentiment in articles on AI and Data Science decreased. A similar trend was observed with Microsoft, which had been heavily investing in ChatGPT. This decline in sentiment could be attributed to initial fears and hesitancy regarding the capabilities of the chatbot, including concerns about its improper use in schools (such as academic dishonesty) and apprehension regarding mass data collection and analysis.

However, starting in January 2023, there was a gradual increase in sentiment. This upward trend coincided with the confirmation of **Microsoft's $10 Billion investment** in ChatGPT and the **growth in popularity** and usage among the public. It is common for new technologies to initially face hesitancy and skepticism before gaining acceptance.

Sentiment Effects on Others Organizations and Products
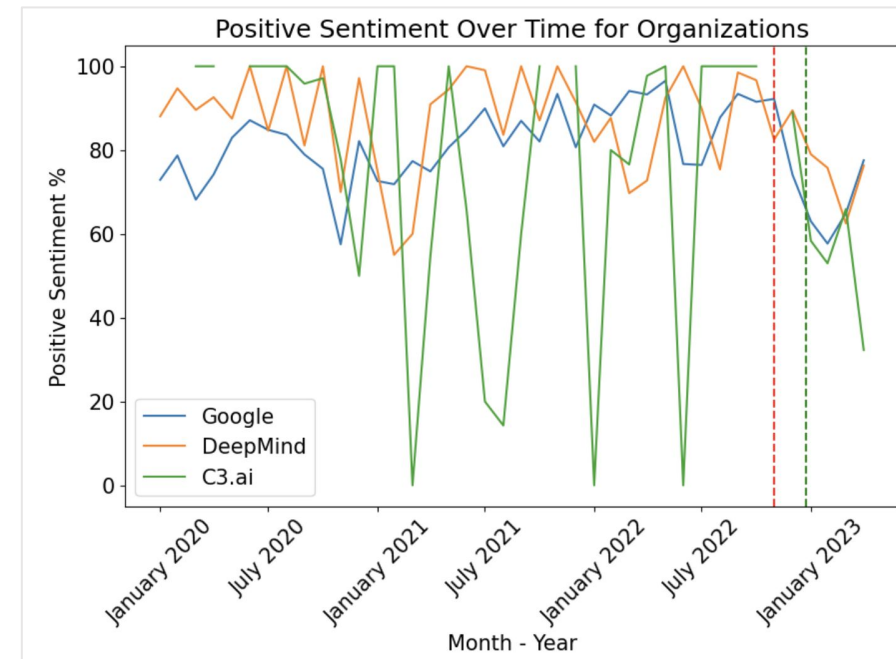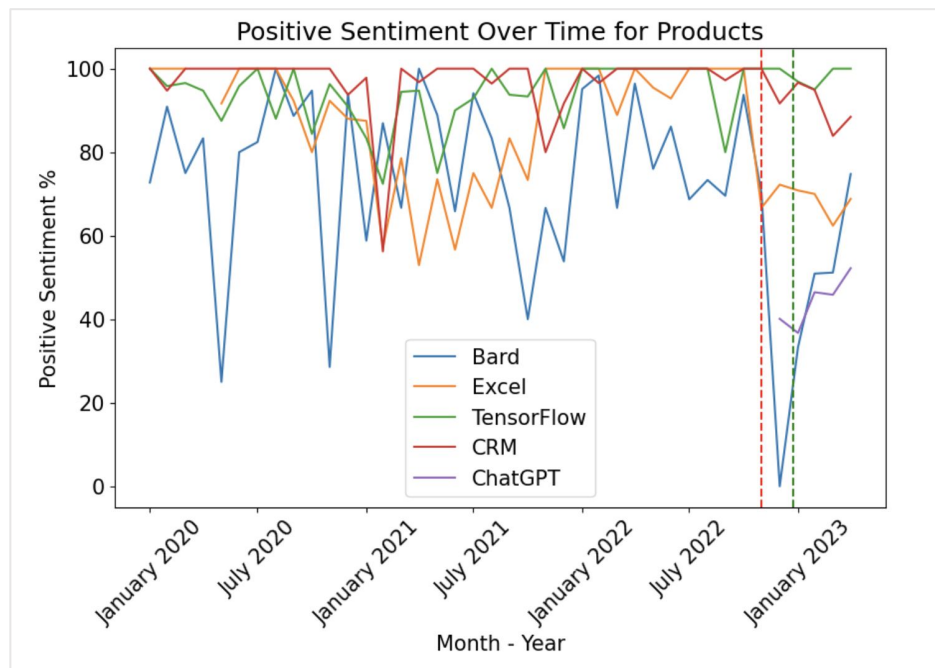
ChatGPT's popularity as the most popular AI tool of 2022 and 2023 could have influenced similar AI tools, such as **Google Bard**, to experience similar sentiment trends. However, established and widely used tools like **Excel** may not have been greatly affected by the release of ChatGPT or other AI advancements.

Furthermore, Google DeepMind, focused on AI research, and CRM IBM Software, focused on customer relationship management, may not have experienced significant changes in sentiment as their activities primarily target the scientific/ business community rather than the general public.
Focused AI tools may have higher sentiments since they are designed to address specific areas and improve specific tasks.

CRM IBM Software: streamlines sales, marketing, and customer service teams → automates manual tasks such as tracking customer information and managing sales.
Google DeepMind: leverages AI technologies to automate and optimize various research processes → frees up researchers' time, allowing them to focus on more complex and creative aspects of their work.

Finding Discussion Surrounding AI Popularity with Text Summarization

Rejecting AI

"Artificial Intelligence (AI) makers … are **hurriedly and desperately** aiming to bring their generative AI apps into the marketplace."

Article Title: Recent ChatGPT And Bard Predicament Raises Thorny Questions About Whether Using One AI To Train A Competing AI Can Be Fair And Square, Says AI Ethics And AI Law
Published by: Central Point News

Accepting AI

" AI and AI-driven virtual beings at a stage like LEAP [have] become a global window into the state of emerging technology and the **most exciting advancements** the industry is experiencing."

Article Title: Sensorium to Lead Conversation on AI Virtual Beings at LEAP 2023
Published by: Yahoo Finance

These statements present two sides of the AI landscape and its recent advancements.

The competitive and fast-paced nature of AI development raises concerns regarding **ethical implications**. The rush to bring AI products to market swiftly may lead to inadequate testing and the introduction of biased algorithms, which can result in privacy breaches or unreliable AI systems that harm individuals or perpetuate societal inequalities.

On the other hand, the exciting potential of AI applications can **positively impact** various aspects of our lives. AI has the capacity to revolutionize industries, enhance productivity, improve healthcare outcomes, and enable personalized experiences.

THE UNIVERSITY OF CHICAGO

Positive and Negative Discussions Surrounding AI
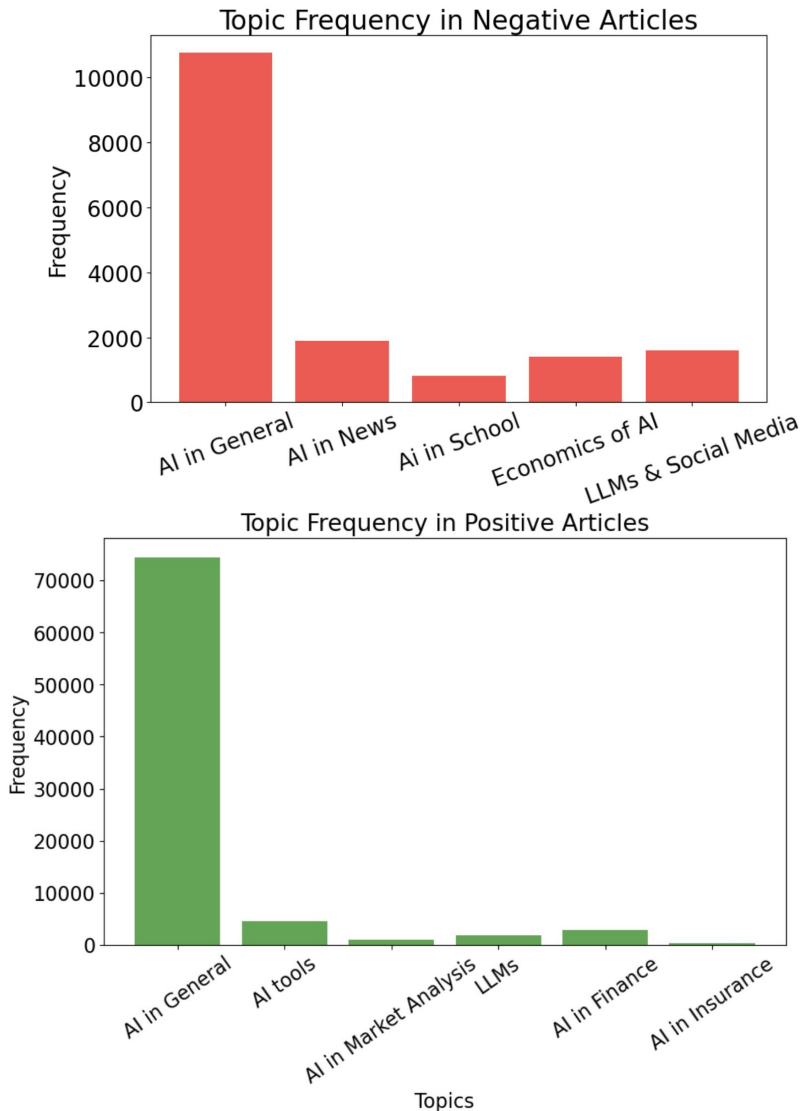
There are political and economic worries surrounding AI, which stem from many political leaders not understanding the capabilities and limitations of AI.

Similarly, fields such as **News**, **School**, and **Social Media** have not embraced AI due to their natural emphasis on creativity rather than data-driven decision-making. School teachers express concerns about academic dishonesty. Similarly, the News and Media industry harbors worries about the <u>lack of transparency</u> in AI algorithms and the <u>diminishment of human judgment</u> in news reporting.

There is continuous development and discussion around AI tools and language models (LLMs). These discussions often involve those who are knowledgeable about technology and less apprehensive about AI advancements.

Similarly, industries such as **Market Analysis**, **Finance**, and **Insurance** have accepted AI technologies since they <u>extensively utilize data</u> in their decision-making processes. They are more likely to create data science roles and remove mundane roles like data entry, inventory management, and customer support.

Note: In the graphs, Topic 0 'General AI and Data Science' was omitted in order to focus on industries affected.

Conclusions

Key Findings

- Industries that benefit from the human touch exhibit more negative sentiment and do not want the potential loss of human sentiment, emotion, creativity, and judgment.
  - Schools (teachers), News (reporters), Media (content manager), Art (artists)
- Industries with prior experience in handling data exhibit more positive sentiment towards AI.
  - Business (consultants), Finance (Stock Trader), Healthcare (Researcher), Insurance (Reporting Developer)
- Focused AI tools exhibit higher sentiments since they are designed to address specific areas and improve specific tasks.
  - Improve: Analyzing Patient Data, Classifying Customer Relationships, Improving Marketing Techniques
- Ethical concerns, privacy issues, and the potential impact on political and economic dynamics contribute to negative sentiment. Lack of existing regulations for AI further amplifies concerns.

Limitations & Future Work

- Limited Computational Power
  - Test more advanced and complicated language models (LLMs) that require greater computational resources.
  - Test and experiment with a wider range of hyperparameters for Latent Dirichlet Allocation (LDA) to enhance topic modeling accuracy with greater computational resources.
- Sentiment Analysis
  - Customize sentiment analysis models further by leveraging a more relevant and specific dataset (Ex: News Articles Discussing Technology).
  - Increase manual labeling efforts to create a larger annotated dataset for training sentiment analysis model.
- Text Summarization
  - Explore additional text summarization techniques to improve quality of summarization.

Applications & Recommendations

Recommendations for AI Accepting Industries: Boost Data Science Utilization

- Reduce Mundane Jobs and Manual Labor:
  - Ex - Customer Support: AI-powered chatbots and virtual assistants handle routine inquiries, freeing up time to focus on more complex customer inquiries.
  - Ex - Manufacturing: AI-powered robotic systems automate assembly line tasks, reducing manual labor in repetitive processes.
- Develop external or internal AI tools:
  - Ex - Healthcare: Development of AI-based diagnostic tools to aid doctors in analyzing medical images and making more accurate diagnoses.
- Mandatory Classes on Relevant AI Tools for Employees:
  - Ex - Consulting: Mandatory workshops on AI tools for consultants, enabling them to leverage AI capabilities in their client engagements.

Recommendations for AI Hesitant Industries: Address Concerns

- Regulatory Frameworks:
  - Support the development of regulatory frameworks that ensure responsible AI use, addressing concerns related to privacy, ethics, and fairness.
- Transparent Practices:
  - Emphasize the importance of transparency by clearly communicating how AI models work, the data used, and the decision-making process behind AI-driven systems.
- Education and Awareness:
  - Provide comprehensive workshops to demystify AI and data science, informing industries about the benefits, limitations, and ethical considerations of AI technologies.
- Continuous Evaluation and Improvement:
  - Have continuous evaluation and accountability of AI systems to address concerns and ensure alignment with industry requirements.

THANK YOU

code: https://github.com/AnkitaNambiar/nlp-UnveilingTheInfluenceOfAiInIndustries

THE UNIVERSITY OF CHICAGO