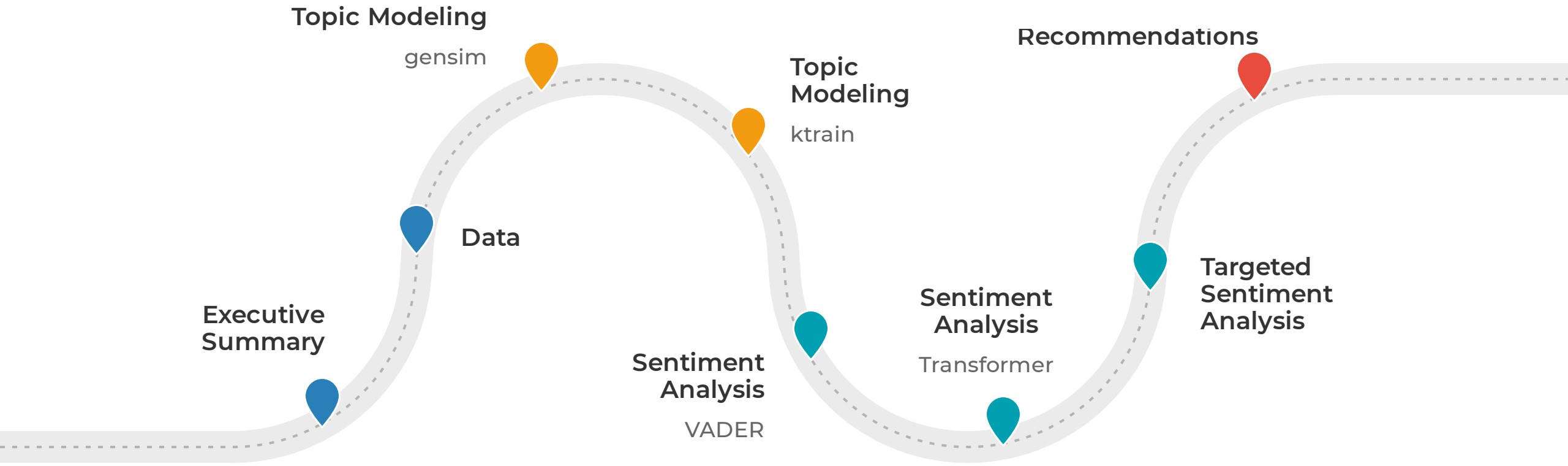
Influence of AI, Data Science & Machine Learning on Industry and their utilization

Isabella Xue

Outline


Executive Summary
Data
Topic Modeling
gensim
Sentiment Analysis
VADER
Topic Modeling
ktrain
Sentiment Analysis
Transformer
Recommendations
Targeted Sentiment Analysis

Executive Summary

In an era of rapid technological evolution, this analysis uncovers pivotal insights at the intersection of AI technology, sentiment analysis, and job opportunities. By focusing on results, it illuminates actionable strategies for capitalizing on AI's transformative potential.

Enhanced Topic Modeling

Employing advanced techniques improved the quality and clarity of identified AI-related topics. This enhanced understanding supports better decision-making and strategic planning.

Sentiment Trends and Influencers

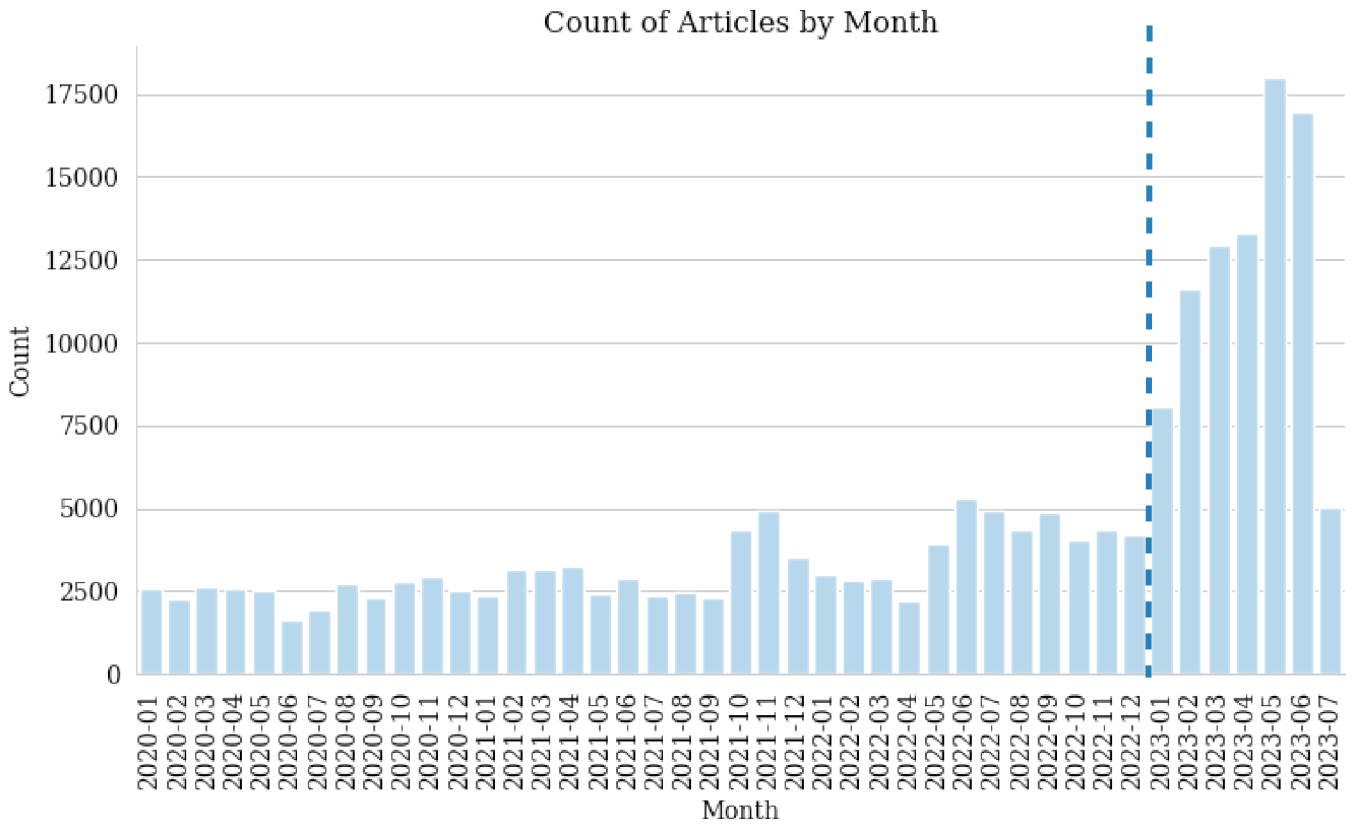
Sentiment analysis uncovered shifting public attitudes toward AI since 2020. Influential individuals, including Sam Altman and prominent politicians, influence sentiment dynamics, highlighting avenues for focused engagement.

AI Integration into Job Roles

Targeted identification of roles ripe for AI integration, such as IT Project Managers and Business Development Managers, underscores the potential for workforce optimization and job transformation.

Prominent Organizations in AI Discourse

The frequent appearance of CHATGPT, Google, Microsoft, Android, and IBM across positive and negative sentiment articles underscores their pivotal roles in shaping AI discussions



News articles regarding AI, DS, and ML have seen a sharp spike since the beginning of 2023.

Data Exploration

- 199,208 news articles
- All values in English
- '*url*', '*date*', '*language*', '*title*', '*text* columns
- No null value in any columns
- Articles range from **January 1st, 2020** to **July 15th, 2023**

Data Processing

- Combine 'title' and 'text' columns
- Remove newlines \n, tabs \t, and other unwanted characters \s
- Remove punctuation
- Remove URLs
- Convert to lowercase
- Lemmatization
- Tokenization

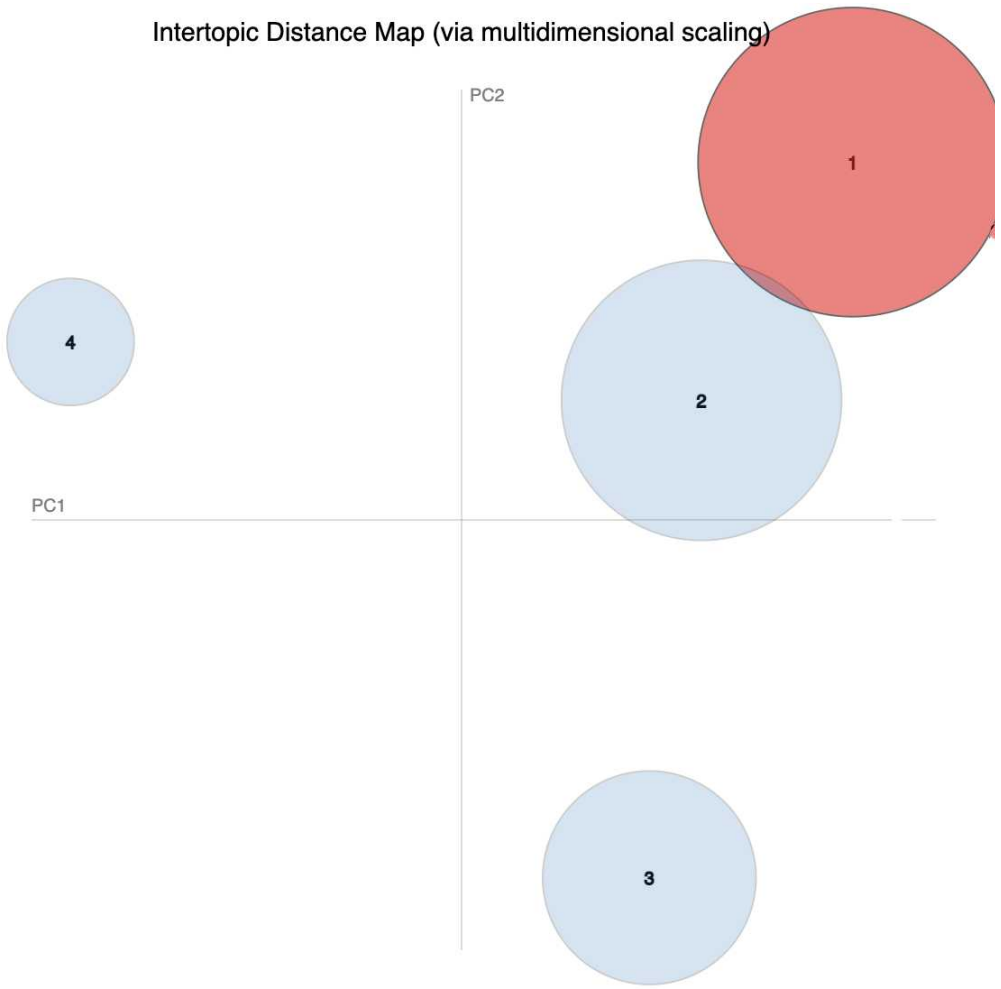
Data: Data Exploration and Processing

Data: Data Filtering

Keyword filtering:

- Find and remove content that does not relate to keyword searches, discarding any materials that are not related to the keywords specified.
- Keywords for filtering include: '[Aa]rtificial [Ii]ntelligence', '[Mm]achine [Ll]earning', '[Dd]eep [Ll]earning', '[Nn]eural [Nn]etworks', '[[The text] was rewritten in active voice.Cc]onvolutional [Nn]eural [Nn]etworks', '[Rr]ecurrent [Nn]eural [Nn]etworks', '[Nn]atural [Ll]anguage [Pp]rocessing', '[Dd]ata [Mm]ining', '[Bb]ig [Dd]ata', '[Dd]ata [Ss]cience', '[Dd]ata [Aa]nalytics' etc.
- Approximately 21% of the articles are excluded from the results
- For optimizing computational speed, randomly choose 70% of data for further analysis

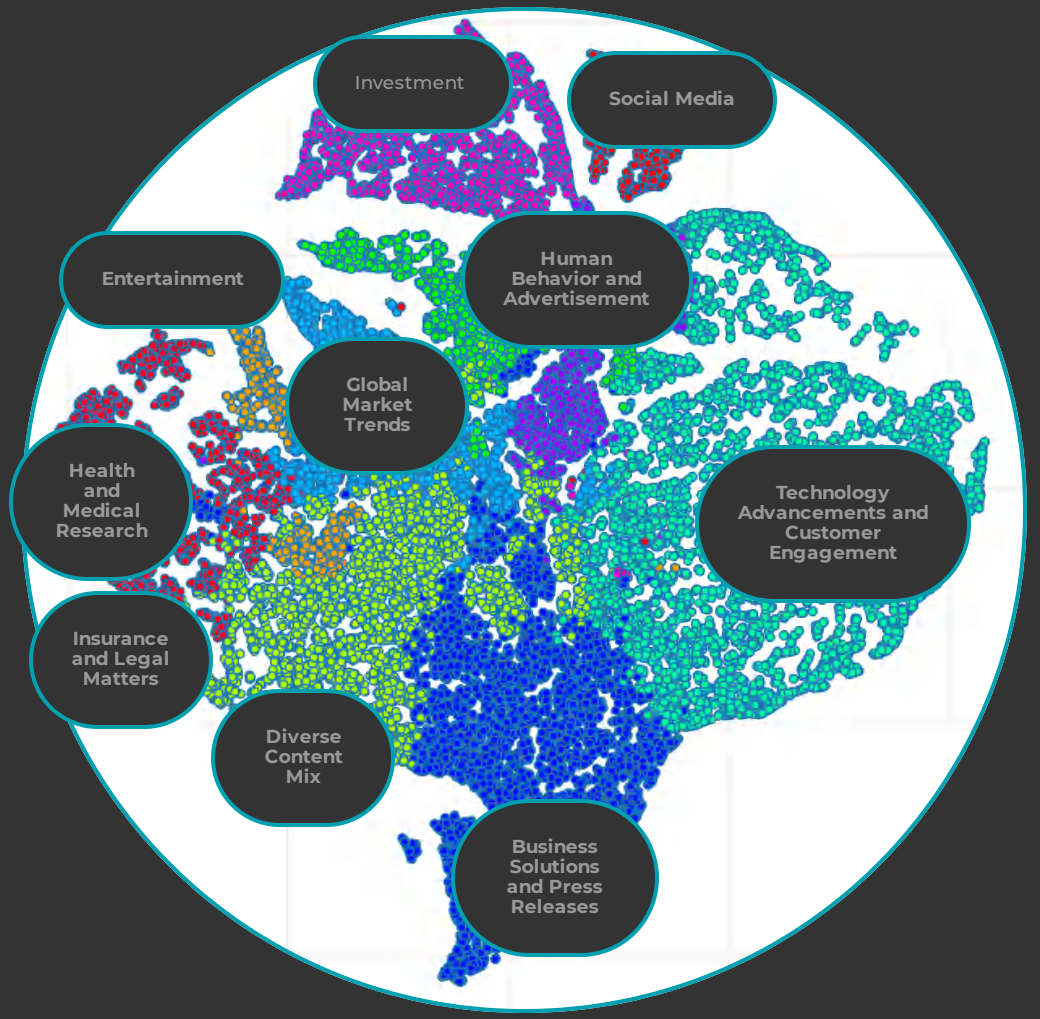
Topic Modeling: LDA gensim



- Multicore LDA
- Performed hyperparameter tuning to find the optimal number of topics, alpha, and beta values that yielded the highest coherence score
  - Topics=4, Alpha='asymmetric', Beta='auto', with a coherence score of 0.479337
- Extracted top words for each of the 4 topics:
  - Topic 0: Market, Intelligence, Artificial, Report, Global
  - Topic 1: AI, Data, Technology, Company, New
  - Topic 2: AI, News, U, New, 2023
  - Topic 3: News, Business, Service, Product, Technology
- Observed that results were not as clean and interpretable as desired. Topics contained less distinct terms.

Topic Modeling: LDA ktrain

- To enhance results, ktrain was chosen as an alternative approach.
- Topics generated:
  - Topic 6: Discussions related to technology tools, digital aspects, customer interactions, security, and the cloud
  - Topic 7: News related to business solutions, platforms, customer engagement, press releases, and digital statements
  - Topic 1: Content about various subjects such as videos, weather, sports, and app-related discussions
  - Topic 3: Content related to health, medical research, machine learning, and studies involving patients and students
  - Topic 0: Analysis of global markets, growth forecasts, key trends, and market players' activities
  - Topic 5: Discussions centered around human behavior, images, work, thoughts, and advertisements
  - Topic 8: Investment-related content including stock markets, funds, prices, buying decisions, and finance
  - Topic 9: Conversations involving social media platforms, community interactions, and local search subscriptions
  - Topic 4: Entertainment releases, consumer resources, general overviews, health, and financial aspects
  - Topic 2: Topics related to insurance, legal matters, African regions, and news updates from various places.



Top Job Candidates for AI Integration and Why

A comprehensive list of job roles across various industries was compiled. These roles encompassed fields such as Human Resources, Marketing, Sales, Operations, IT, Healthcare, Education, Engineering, and more.

The processed text of each news article was transformed to lowercase, and a check was conducted to identify job-related keywords from the list in the article's text.

Top 10 candidates and Why:

Data Entry Clerk: AI can swiftly process and input data, reducing manual data entry tasks.

IT Support Specialist: AI chatbots can troubleshoot technical issues, enhancing customer support.

Business Development Manager: AI can analyze market trends to identify growth opportunities.

Customer Service Representative: AI-powered chatbots can provide instant responses to customer queries.

Data Scientist: AI can quickly analyze vast datasets for insights and predictions.

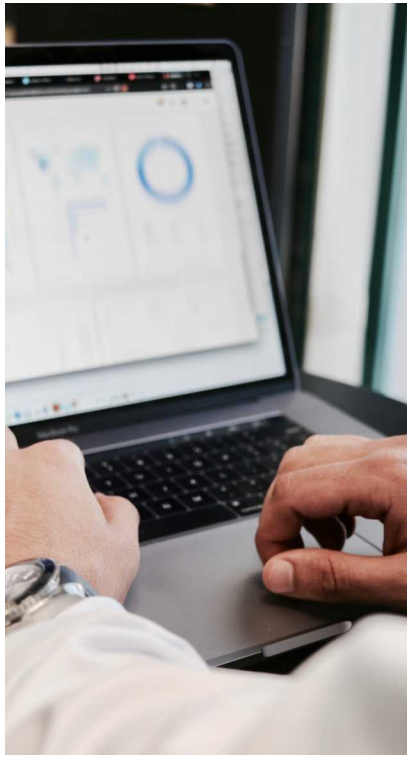
HR Coordinator: AI can streamline candidate screening and shortlisting processes.

Content Writer: AI tools can assist in generating content ideas and maintaining consistency.

Social Media Manager: AI can suggest optimal posting times and content for engagement.

Product Manager: AI can analyze market demand and user feedback for product development.

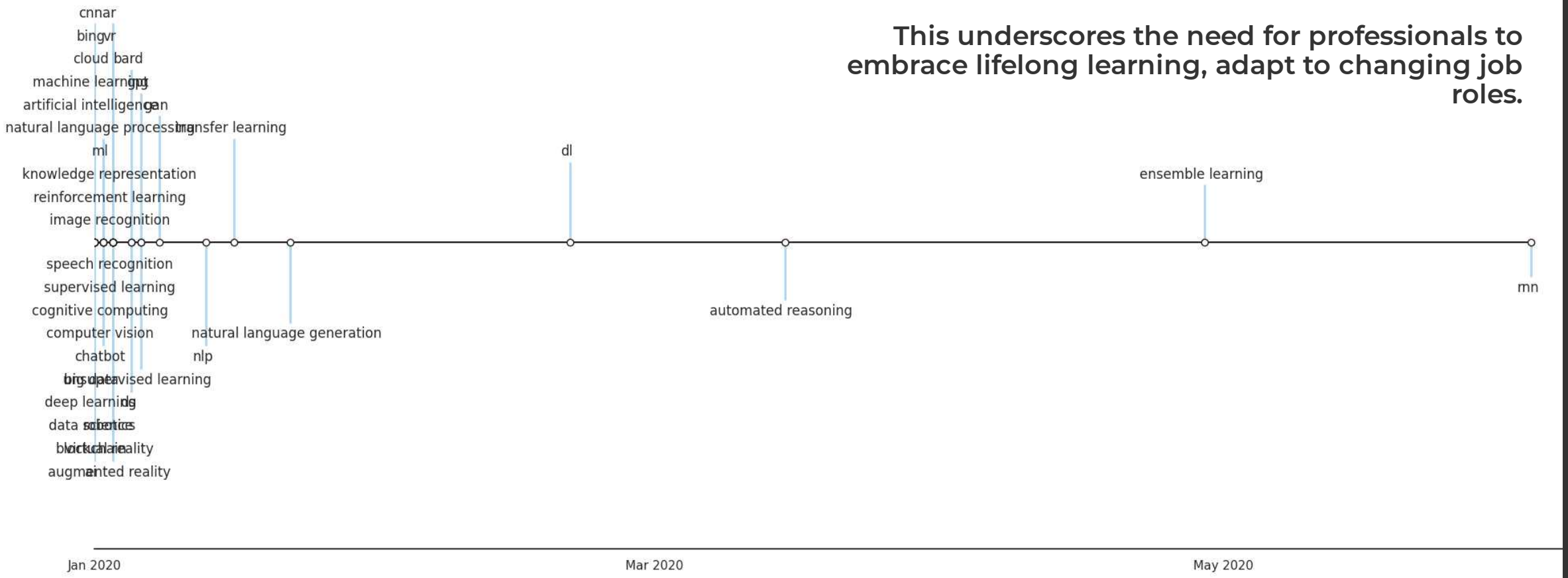
Public Relations Manager: AI sentiment analysis aids in gauging public perception.



Explore innovative tech and AI advancements

All of the technologies emerged before June 2020. This indicates a rapid adoption and integration of these technologies into various industries and sectors.

This underscores the need for professionals to embrace lifelong learning, adapt to changing job roles.



Sentiment Analysis

Methodology

VADER

- Utilized VADER (Valence Aware Dictionary and sEntiment Reasoner) for sentiment analysis.
- Employed SentimentIntensityAnalyzer from NLTK library to assess sentiment in processed text.
- Assigned sentiment scores including positive, negative, and neutral sentiments, along with a compound score.
- Labeled text as 'positive' (compound score ≥ 0.05), 'negative' (compound score ≤ -0.05), or 'neutral' (-0.05 < compound score < 0.05).

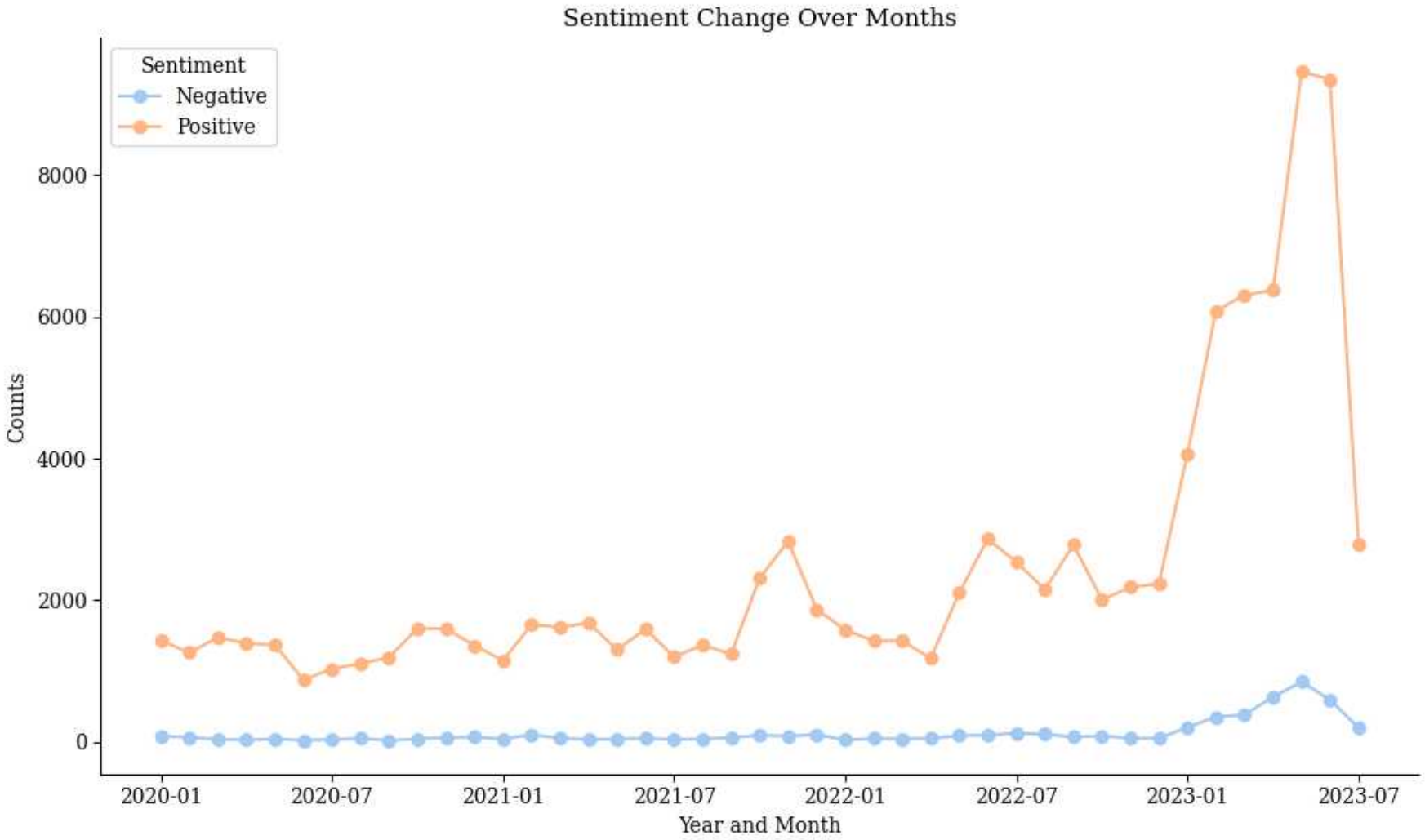
Transformer

- Deployed fine-tuned version of distilroberta-base on the financial_phrasebank dataset (*https://tinyurl.com/mr39ywt6*)
- Employed the AutoTokenizer from the Hugging Face Transformers library to tokenize and prepare text for the transformer model. Utilized the AutoModelForSequenceClassification to perform inference on the tokenized input.
- Calculated sentiment probabilities for each sentiment class using the softmax function on the output logits. The probabilities were used to determine the sentiment label.
- Broke down the dataset into four separate dataframes based on years (2020, 2021, 2022, and a sample from 2023) to ensure computational efficiency.
- This transformer-based approach allowed for fine-grained sentiment analysis across financial news articles, enabling the classification of sentiments into predefined categories based on the content's emotional tone.

Sentiment Analysis

Sentiment Overtime

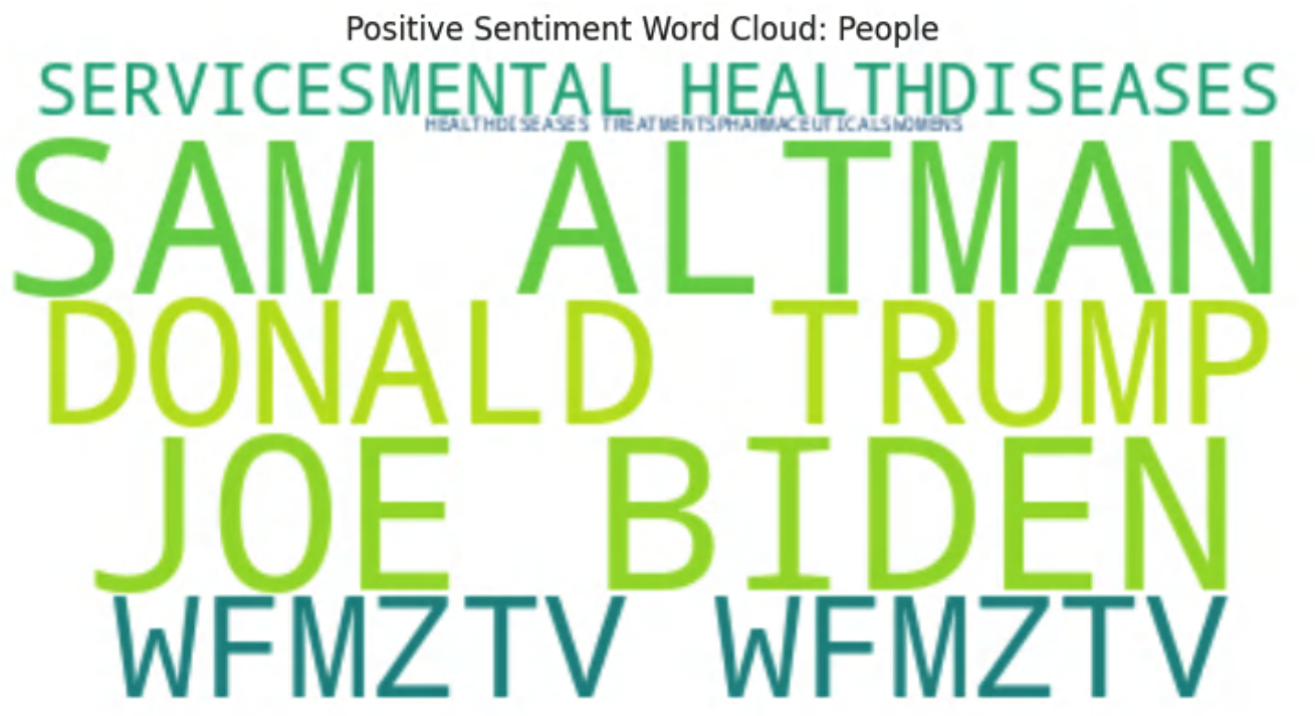
A steady trend of negative sentiment has been observed since 2020, however the number of positive news articles relating to AI saw a sharp increase in early 2023.

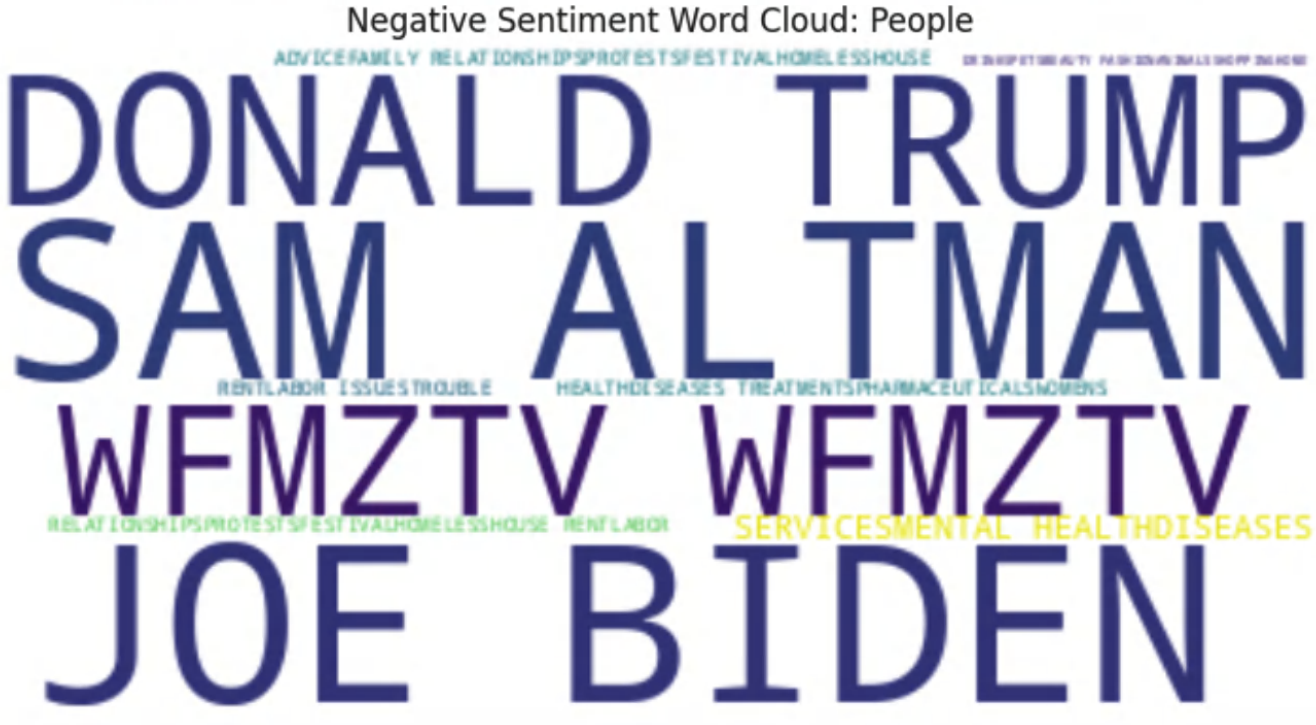


Sentiment Analysis

People

Sam Altman, Donald Trump, and Joe Biden appear prominently in news articles with positive and negative sentiment alike according to the wordclouds.





Sentiment Analysis

Organization

No clear conclusion can be drawn based on the wordcloud since there is not enough proof. Nevertheless, the comparison of positive and negative sentiment wordclouds reveals that these major organizations are present to a similar extent in related news articles.

Positive Sentiment Word Cloud: Organizations

Negative Sentiment Word Cloud: Organizations

Actionable Recommendations

Automate routine tasks

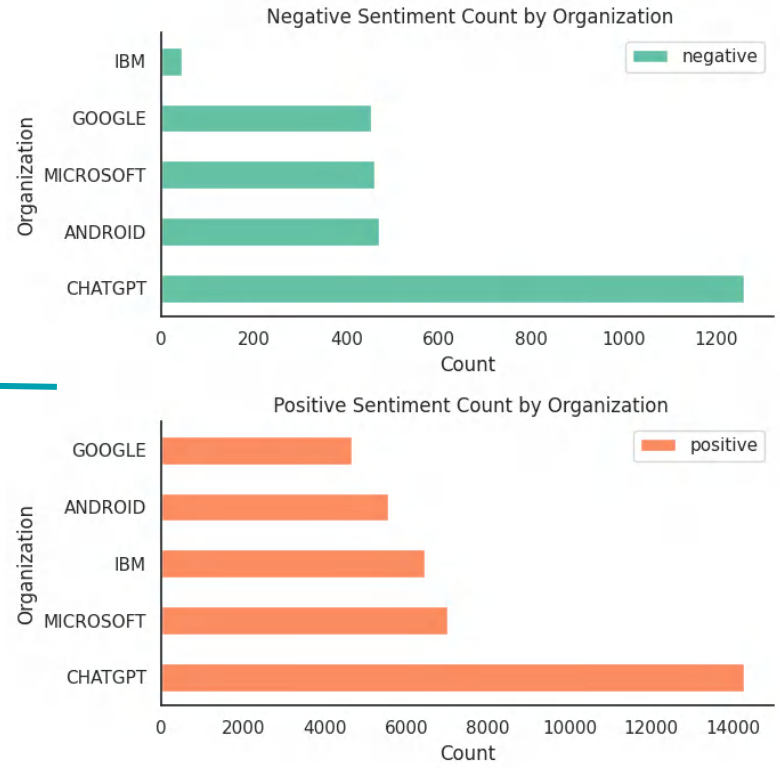
- Utilize AI technologies, such as chatbots and process automation tools, to streamline routine tasks identified in job roles like Customer Service Representatives and Data Entry Clerks
- By automating repetitive tasks, employees can focus on more strategic and value-added activities

AI Impact Assessment

- Conduct regular assessments of AI's societal impact, including job displacement, skill demands, and societal transformations
- Use insights from these assessments to inform policy decisions, workforce training programs, and strategic planning

Industry Collaboration and Innovation

- Foster collaboration between organizations, academic institutions, and government bodies to drive AI advancements and address societal challenges
- Establish industry partnerships to pool resources and expertise for AI research and development



Github link:

https://github.com/isabellaxue/news_articles_nlp