RISE RESEARCH INNOVATION SCHOLARSHIP ENTREPRENEURSHIP

Mentor Name: John Rachlin
Subject/Category: Data Science
Class: DS3500
Group #: 30

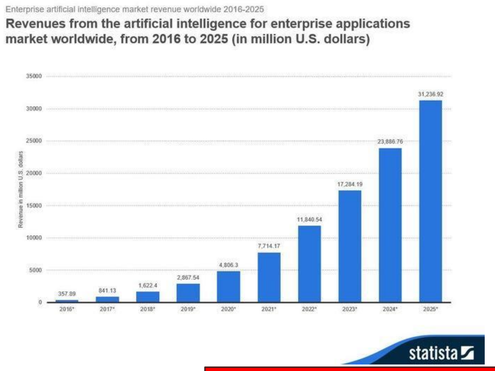
Northeastern University

AI-gniting Insights: Where NLP meets WSJ

Daniel Veretenov, Vivian Li, Ryan Wu

BACKGROUND

With the rise of Artificial Intelligence year after year, it seems its growth is almost limitless and is destined to only become further integrated into our everyday lives, expected to expand **37%** every year from **2023 to 2030.** However, AI technology seems to be both a great tool and a looming threat for many. Through utilizing an NLP framework, we hope to analyze a set of news articles sourced from the Wall Street Journal related to AI to hopefully draw conclusions on the perception, effects, and implications it has on our future.



BACKGROUND

- text preprocessing file:
  - Preprocessed the data by removing any punctuations & white spaces
  - count the occurrences of unique words and calculate average word length
- npl library:
  - loaded all the text into preprocessing & store data about each file
  - load the stop word into a file
  - wrapper methods to generate visualizations using Visualizations class]
- visualizations:
  - wordcount_sankey: creates a Sankey diagram visualizing word frequencies
  - word_cloud_subplots: Generates subplots of word clouds for each text entry
  - word_frequency_overlay: Produces a bar plot comparing the top 15% word frequencies across different text entries
- main:
  - instantiates an NLP library object, loads stop words and text files, and generates visualizations

PRODUCTS AND CONCLUSIONS

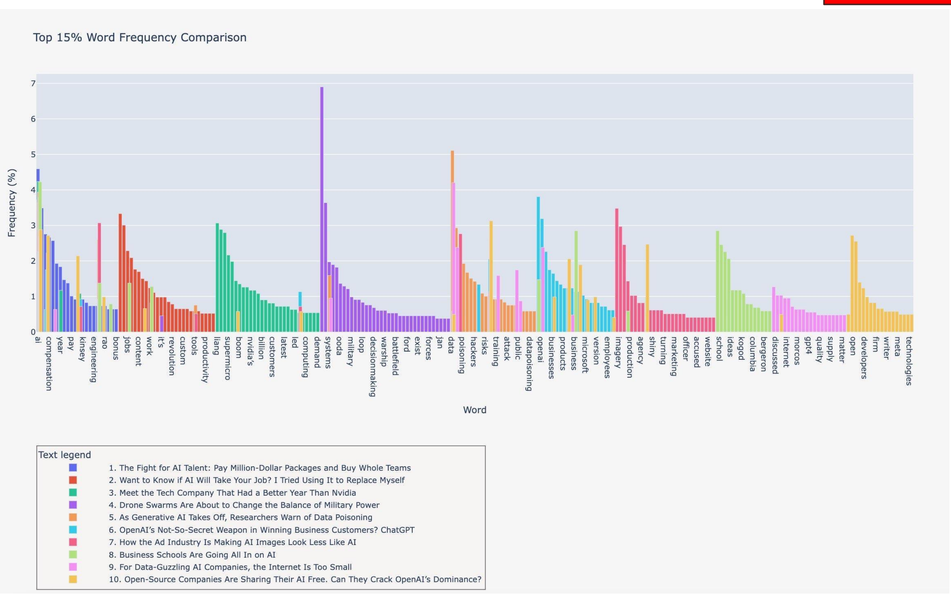
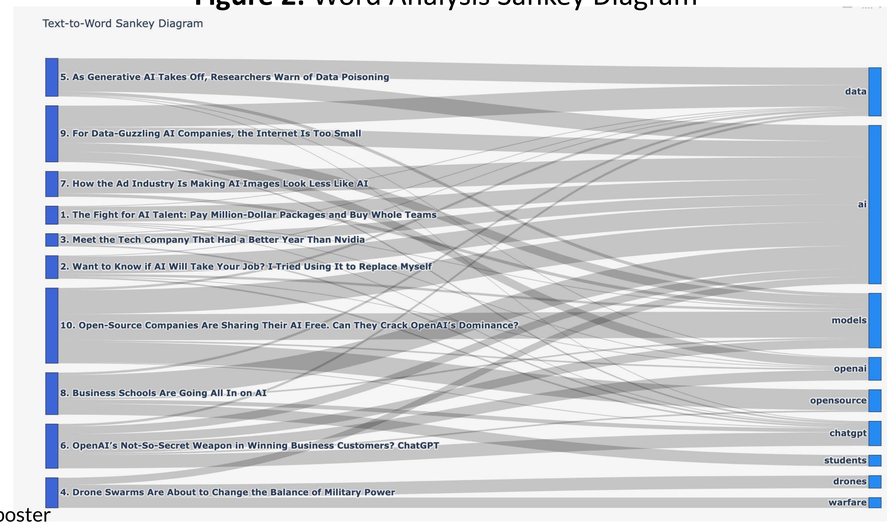


Figure 1: Interactive Word Count Dashboard

Figure 3: Word Cloud Subplot Diagram

Figure 2: Word Analysis Sankey Diagram



With the visualizations above, observations regarding the impact of AI on employment and job relations can be made. Figures 1 and 3 illustrate that across these 10 articles, mentions of compensation, jobs, and businesses are prevalent. Despite covering different subjects, industries, or methodologies, each article relates back to the utilization of AI. Additionally, an interesting contrast emerges between articles discussing generative AI, with one portraying negative connotations ("poisoning, risks, attack") and another highlighting positive aspects ("shiny, agency"), thereby underscoring the controversy surrounding AI's role in creating art and other forms of media. Figure 2 emphasizes the prevalence of OpenAI and ChatGPT, as they are the only names present in a sea of general terms within the Sankey diagram. OpenAI's substantial influence on the broader AI landscape elucidates its significant presence among these articles.

In conclusion, the analysis of AI news articles from the Wall Street Journal highlights the widespread impact of AI across various sectors, including compensation, job relations, businesses, and industries. Despite discussing diverse subjects, each article ties back to AI utilization. Notably, conflicting perspectives on generative AI reveal the ongoing debate surrounding its adoption. Additionally, the dominance of key players like OpenAI and ChatGPT underscores their influence in shaping AI discourse. Overall, the analysis emphasizes the multifaceted nature of AI discussions, necessitating comprehensive approaches to address its implications effectively.

Author Contributions:
Daniel Veretenov: npl codes, sankey & wordcloud visualizations | Vivian Li: sankey & wordcloud visualization editor, bar chart visualization and poster | Ryan Wu: bar chart visualization and poster