**Data Science Project Proposal: Sentiment Analysis on AI and the Job Market**

* By Siling Song

Artificial Intelligence is now a heated topic in advanced economies. Its influence on job prospects is an even more intense debate. Is the concern that AI will inexorably wipe out jobs valid, or is it a consequence of media exaggeration under the belief that negative news absorbs most attention from the audience? Is the sentiment prevailing in the workforce and spreading on the media consistent with the findings discovered by scholars in the field? Did they change across the years? This project would seek to answer these questions using the technique of sentiment analysis.

**Central Research Questions:**

1. Sentiment analysis of the workforce:

* How does the American workforce view the role of Artificial Intelligence in the job market?
* Is it pessimistic or optimistic?
* Did it change across years, especially after the launch of Chat GPT?

1. Sentiment analysis of the Media:

* How is the relationship between Artificial Intelligence and jobs portrayed by the media?
* Is it pessimistic or optimistic?
* Did it change across years, especially after the launch of Chat GPT?

1. Sentiment analysis of the scholarly works (Research papers)

* How optimistic are researchers in academia/think tanks on the prospect of work with the accelerated development of Artificial Intelligence?

Sentiment analysis of the hearings.

**Method: Sentiment analysis**

* Scrapping data (title, text, and published date) with keywords ‘AI’ and ‘job’ from three sources:
  + **News articles:** including those from CNN, Fox news, New York Times, The Washington Post etc.
  + **Social Media:** including YouTube platforms (video title & comment area), Instagram, Reddit, Facebook, Twitter,
  + Published Paper. (Could also substitute this with the previous work of literature review that the team has already performed)
* Use NLTK and other tools of natural language processing to examine the sentiment of the scrapped data, categorized into positive, neutral, and negative.
* Use python to visualize the data across the years.
* Potential extension: Semantic analysis to model topics: i.e. what industries/topics is the AI most frequently associated with in media outlets.

**Related Literature:**

This [paper](https://arxiv.org/html/2401.08899v1" \l ":~:text=The%20overall%20discourse%20in%20the,and%2028%25%20positive%20sentiment%20articles" \t "_blank) explored the landscape of Generative AI in terms of Topics, Sentiments, and Spatiotemporal Analysis. They gathered an extensive collection of news articles from 2018 to November 2023. It is a very comprehensive analysis. We can utilize their framework and improve upon their result by utilizing a larger timeframe of data sources (2010-2024) and extending our data sample from news articles to public social media and scholar works.

**Expected Output:**

* Example visualization:
* 

This will give us a direct idea of how people are viewing the influence of AI on the workforce, how that has evolved through time, and how public, media, and scholar opinions may differ.

* Current Progress:

A graph of positive negative and neutral

Description automatically generated

**Next Steps:**

* Break down the data into monthly timeframe and conduct a 2020-2024 analysis
* Break down the news article by category (education, business, politics ect.), conduct sentiment analysis and examine how much they differ.
* News agencies
* Explore and employ more accurate analysis algorithm, comparing the result
* Semantic analysis using BERTopic, especially on news associated with negative sentiments, examine what news articles often cite as concern for AI and job.
* Spatial distribution of the news articles and their relative sentiment analysis.