Project 2: Draft

Econ 1680: Machine Learning, Text Analysis, and Economics

Nadya Tan

April 7, 2024

1 Introduction

The fear that AI and automation would one day take over human jobs is nothing new, but it has become increasingly prevalent in the last couple of years, as developments in the field of machine learning and natural language processing have arguably made some white-collar work redundant (such as with the big rise of generative AI in 2023 (Lynch (2023))). Most job cuts that have been explicitly tied to AI have been in the tech sector, such as IBM's and Klarna's (Constantz (2024)). Simultaneously, the period in the aftermath of the pandemic has been a tumultuous time for the tech sector, with numerous layoffs happening going into 2023 and as the year progressed. Hence, it would be interesting to explore how the news coverage of AI has evolved over this particular time period, to see if public hysteria surrounding new technology could potentially be linked to very public job cuts.

This paper seeks to identify how the sentiment paths of news articles about AI have changed over the recent years, and how this has coincided with the timing of mass tech layoffs. Using a variety of methods, such as transfer-learning methods and lexicon-based methods, to identify sentiment trends in news articles is not new (David Rozado (2022)), and this paper will borrow on some of those methods, but will use it to look at sentiment trends for articles about AI specifically. Moreover, this paper will then utilize topic modelling to identify different clusters of topics within news articles about AI, and see how the prevalence and sentiment of those topics have changed over time. All in all, this paper's main contributions will be in helping us better understand the trends in AI news coverage and technology news coverage as a whole, and how the sentiment of such coverage changes when certain economic developments or shocks occur.

2 Data Sources and Descriptions

To look at the coverage of news articles related to AI over recent years, I used the New York Times' API . This API allows users to search for titles, abstracts and leading paragraphs of articles relating to a certain keyword within a certain date frame. For every month in the years 2020-2023, I searched for the top 50 articles that were related to artificial intelligence. I received a total of 1370 articles, an average of 28-29 articles per month. An overview of

what the API returns can be found in Table 1. A disadvantage to using the API was that it only returns abstracts and leading paragraphs instead of the full body text, but this is still sufficient for the purposes of analyzing overall article sentiment, which should be captured in the abstract that summarizes each article.

In order to analyze trends for articles that related to employment as well as AI, I assembled a list of words related to employment (employment, jobs, job, occupation, work, trade, profession), and created a "mentions employment" dummy variable. Should the abstract contain any of these words, the dummy was set to 1.

Table 1: Sample of Articles Retrieved by NYT API

Date	Title	Paragraph	Abstract
2022-11	What Riding in a Self-Driving Tesla Tells Us About the Future of Autonomy	We took a ride with Tesla's experimental self-driving system. It was by turns intriguing, impressive and scary.	We took a ride with Tesla's experimental self-driving system. It was by turns intriguing, impressive and scary.
2022-12	May I Speak to a Human, Please?	Breana Jones started investing in 2014, building up her retirement savings and putting aside money to buy a house...	Younger investors who are navigating market volatility and trying to save for retirement are finding that roboadvisers lack the personal touch
2023-03	How Siri, Alexa and Google Assistant Lost the A.I. Race	On a rainy Tuesday in San Francisco, Apple executives took the stage in a crowded auditorium to unveil the fifth-generation iPhone...	The virtual assistants had more than a decade to become indispensable. But they were hampered by clunky design and miscalculations, leaving room for chatbots to rise.

In order to access data about tech layoffs, I used a Kaggle dataset that contains the details of tech layoffs up until 2024 that was put together by scraping layoffs.fyi. This database records each round of layoffs, keeping record of the date of the layoff, the company that laid off employees, and in some cases, the number of people affected by layoffs. Figure 1 shows the number of tech layoffs over time in recent years, whereas Figure 2 shows the number of NYT articles related to AI over time during the same period. As visible from these graphs, coverage of AI seems to have increased as the number of tech layoffs increased.

3   Method

In order to analyse the sentiment of these articles, I used the default pre-trained model for sentiment analysis developed by Huggingface (DistilBERT-base-uncased). This model was pre-trained and fine-tuned on the Stanford Sentiment Treebank corpora, and operates on a Transformer architecture, which is a type of deep learning model designed for processing sequential data, such as text. Specifically, it uses the DistilBERT architecture, which is a

Figure 1: Number of Tech Layoffs Over Time
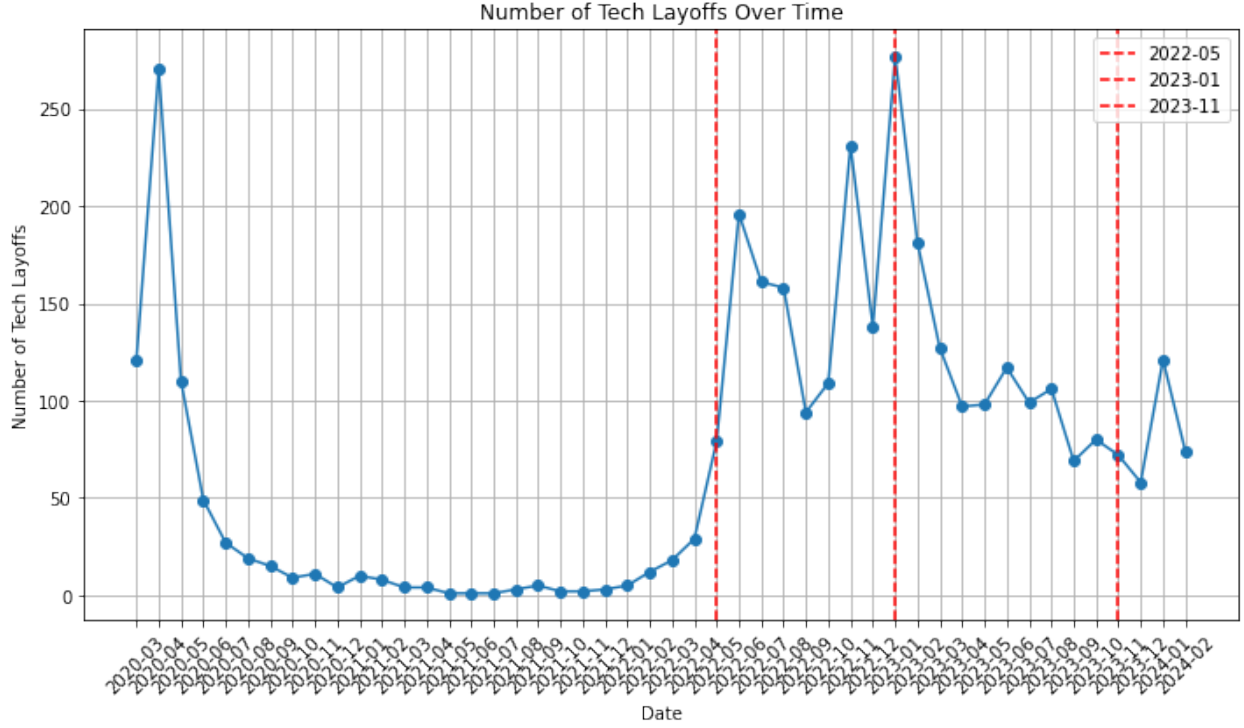
Note: A company-wide announcement of a layoff is counted as one layoff, regardless of the number of individuals impacted. The 3 dashed red lines correspond to the beginning, peak and tail end of tech layoffs

distilled version of the BERT model, making it computationally lighter while retaining much of its performance (Victor Sanh (2019)). The advantage of using transformer models is that they tend to be better at understanding relationships between words and phrases, as opposed to handling each article like a simple bag of words. In addition to classifying an article as either "positive" or "negative", this DistilBERT model also provides a sentiment score from 0-1, which indicates how strong the sentiment in either direction is. I multiplied this score by either 1 (if the article was labelled as positive) or -1 (negative) to get a final sentiment score between -1 to 1, where 0 is neutral.

Topic modelling to identify clusters of articles will be done using BERTopic, a deep-learning based topic modeling method developed by Maarten Grootendorst (Grootendorst (2022)). It leverages BERT embeddings and c-TF-IDF to create dense clusters, allowing for easily interpretable topics whilst at the same time keeping important words in the topic descriptions. I will first remove stopwords in the article abstracts, and restrict the minimum number of documents that I wanted BERTopic to classify into each topic. BERTopic also discards any motifs that were not classified into topics into a discard topic (topic -1), which also helps for interpretability. Topics will be evaluated by eye using sense checks.

Note: My notebook currently shows Logit-Lasso results for words are predictive of sentiment and employment mentions, but I have decided to drop them from the paper for now, as I'm not entirely sure how they contribute to the research questions.

Figure 2: Number of NYT Articles Related to AI Over Time
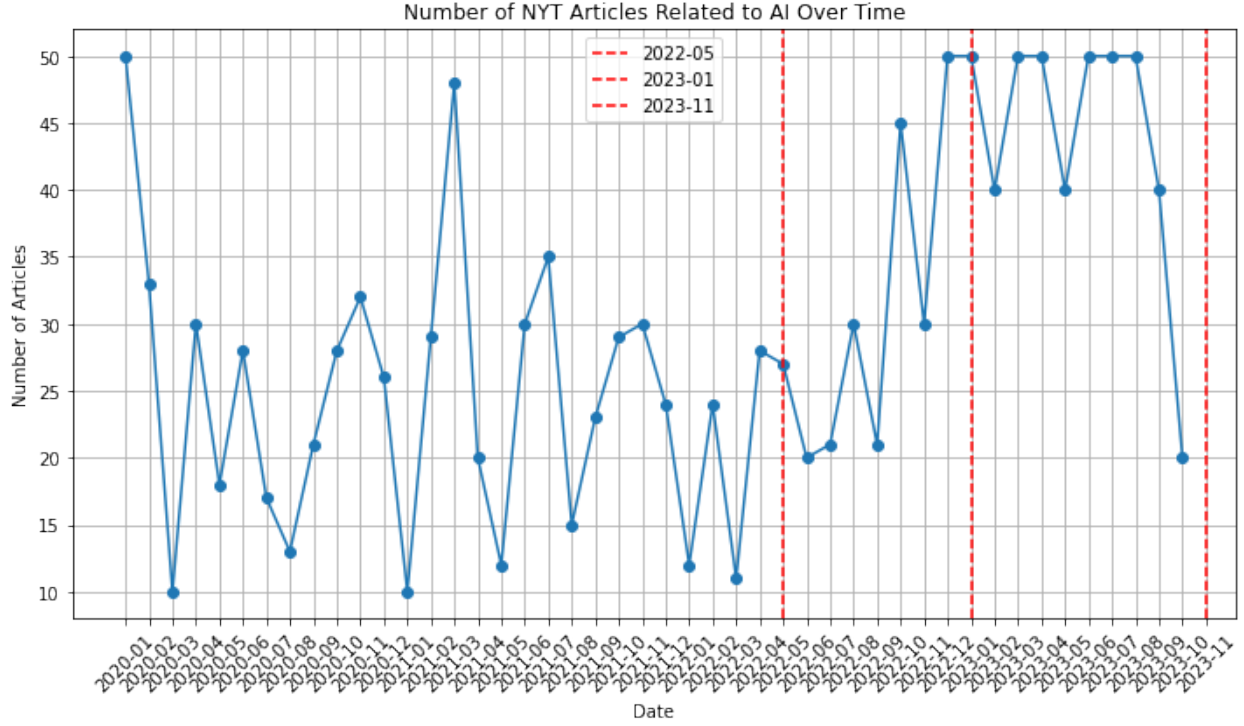
Note: The 3 dashed red lines correspond to the beginning, peak and tail end of tech layoffs

4 Results or Expected Results

Figure 3 shows that the average sentiment score of NYT articles related to AI started to become more negative after the start of the layoffs, reaching a local minimum at the peak of layoffs, and continued being similarly negative even as the number of layoffs started to taper off. Figure 4 demonstrates that this trend could be driven by a disproportionate increase in the number of articles that have negative sentiment during this time.

Besides looking at overall trends, I wanted to look at trends for articles that mention employment, as shown in Figure 5. For this group of articles, we see a more drastic decrease in average sentiment levels between the start of the layoffs and when the layoffs hit their peak.

One way to make sense of these trends is to take a deeper dive into what type of negative articles popped up during the peak layoff period. Table 2 shows a sample of these articles, which shows that whilst some of these articles are directly about layoffs, others are about other aspects of tech companies or general developments in AI.

Topic modelling / clustering analysis to come.

5 Conclusion

Pending.

Figure 3: Average Sentiment Score of NYT Articles Related to AI OVer Time
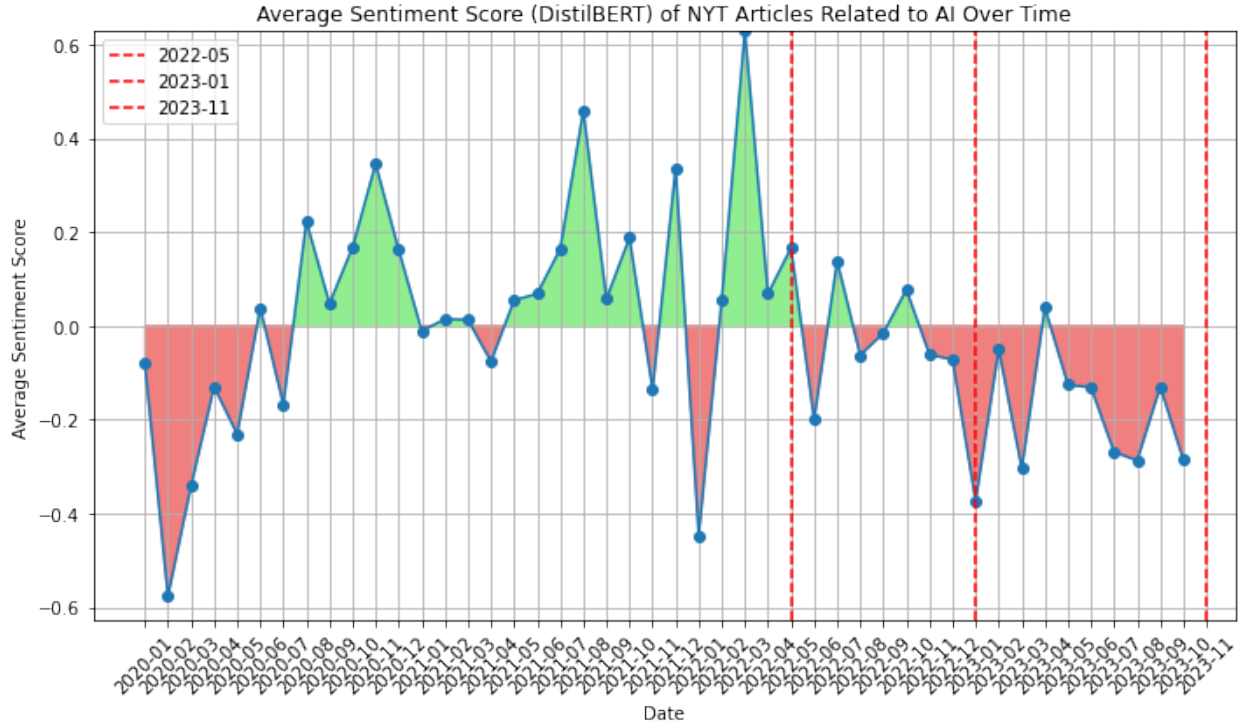
Note: The 3 dashed red lines correspond to the beginning, peak and tail end of tech layoffs

References

Constantz, Jo (2024) "AI Is Driving More Layoffs than Companies Want to Admit," *BNN Bloomberg*.

David Rozado, Jamin Halberstadt, Ruth Hughes (2022) "Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models," *Plos One*.

Grootendorst, Maarten (2022) "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXivLabs*.

Lynch, Shana (2023) "13 Biggest AI Stories of 2023."

Victor Sanh, Julien Chaumond Thomas Wolf, Lysandre Debut (2019) "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *NeurIPS*.

Figure 4: Number of Positive and Negative NYT Articles Related to AI Over Time
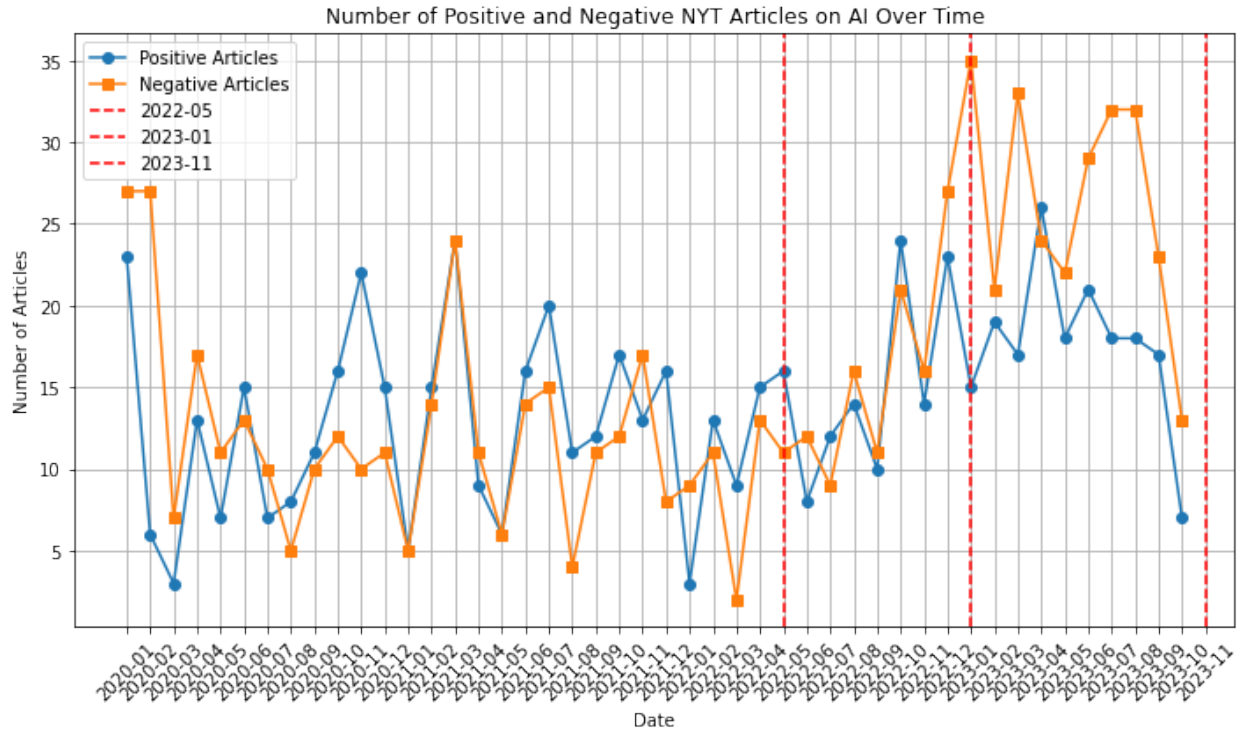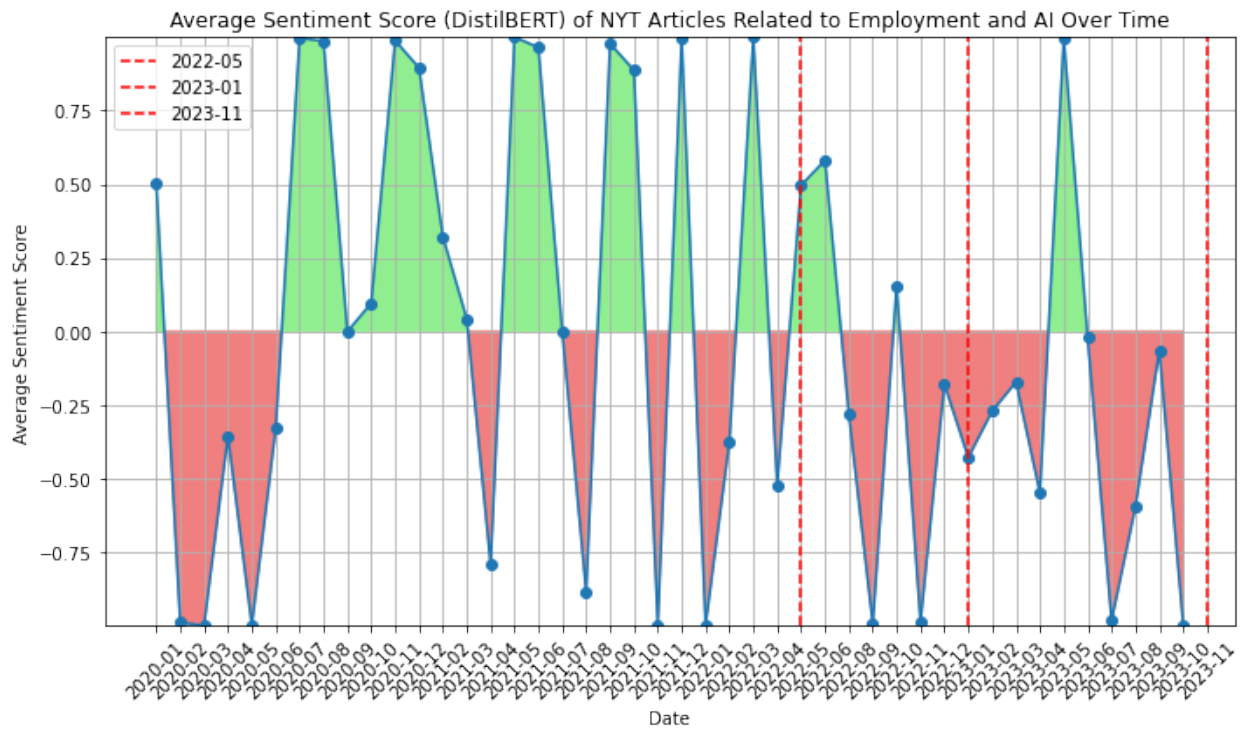

Note: The 3 dashed red lines correspond to the beginning, peak and tail end of tech layoffs

Table 2: Sample of Articles with Negative Sentiment During January 2023 - Peak Layoff Period

Date	Title	Paragraph	Abstract
2023-01	Why Netflix Is Changing the Guard	For 25 years, Reed Hastings was not only Netflix's co-founder, but also its most visible cheerleader and advocate...	Reed Hastings's move to become executive chairman may be a sign that the streaming giant thinks the worst may be behind it.
2023-01	Elon Musk's Appetite for Destruction	Early on, the software had the regrettable habit of hitting police cruisers. No one knew why, though...	A wave of lawsuits argue that Tesla's self-driving software is dangerously over-hyped. What can its blind spots teach us about the company's erratic C.E.O.?
2023-01	Microsoft Revenue Up 2 Percent, but Profit Drops 12 Percent	Microsoft on Tuesday reported its slowest growth in six years and cautioned that a broader slump ...	The company, which announced plans to lay off 10,000 workers last week, had warned that was facing a significant slowdown in the growth of its sales

Figure 5: Number of NYT Articles Related to Employment and AI Over Time

Note: The 3 dashed red lines correspond to the beginning, peak and tail end of tech layoffs