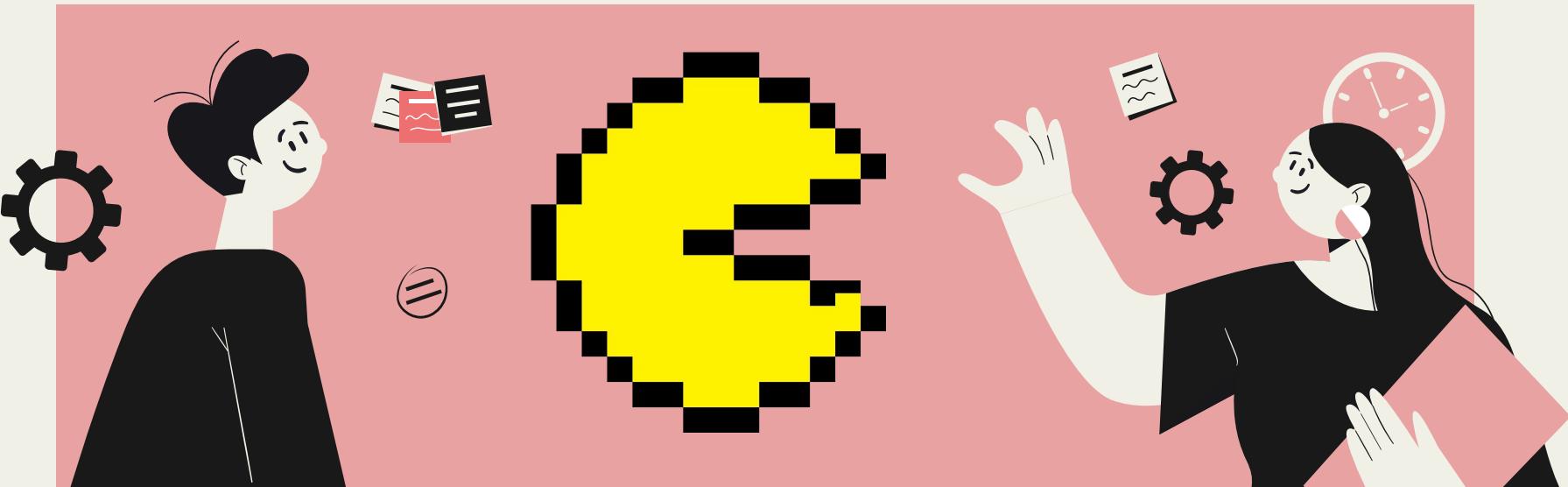


Pixel Pandits

Our Future with AI: Utopian or Dystopian?

Saksham, Pulin

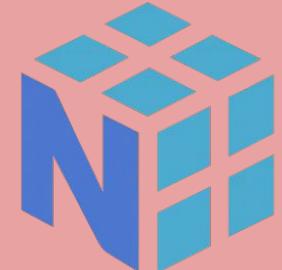
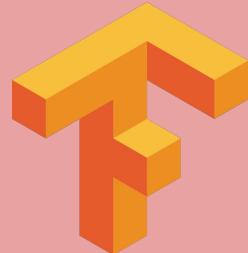
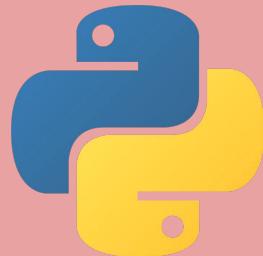
Naman, Dhruv



Introduction

We decided to use the 10K dataset for our submission, as we believed we could deliver insights better on more data.

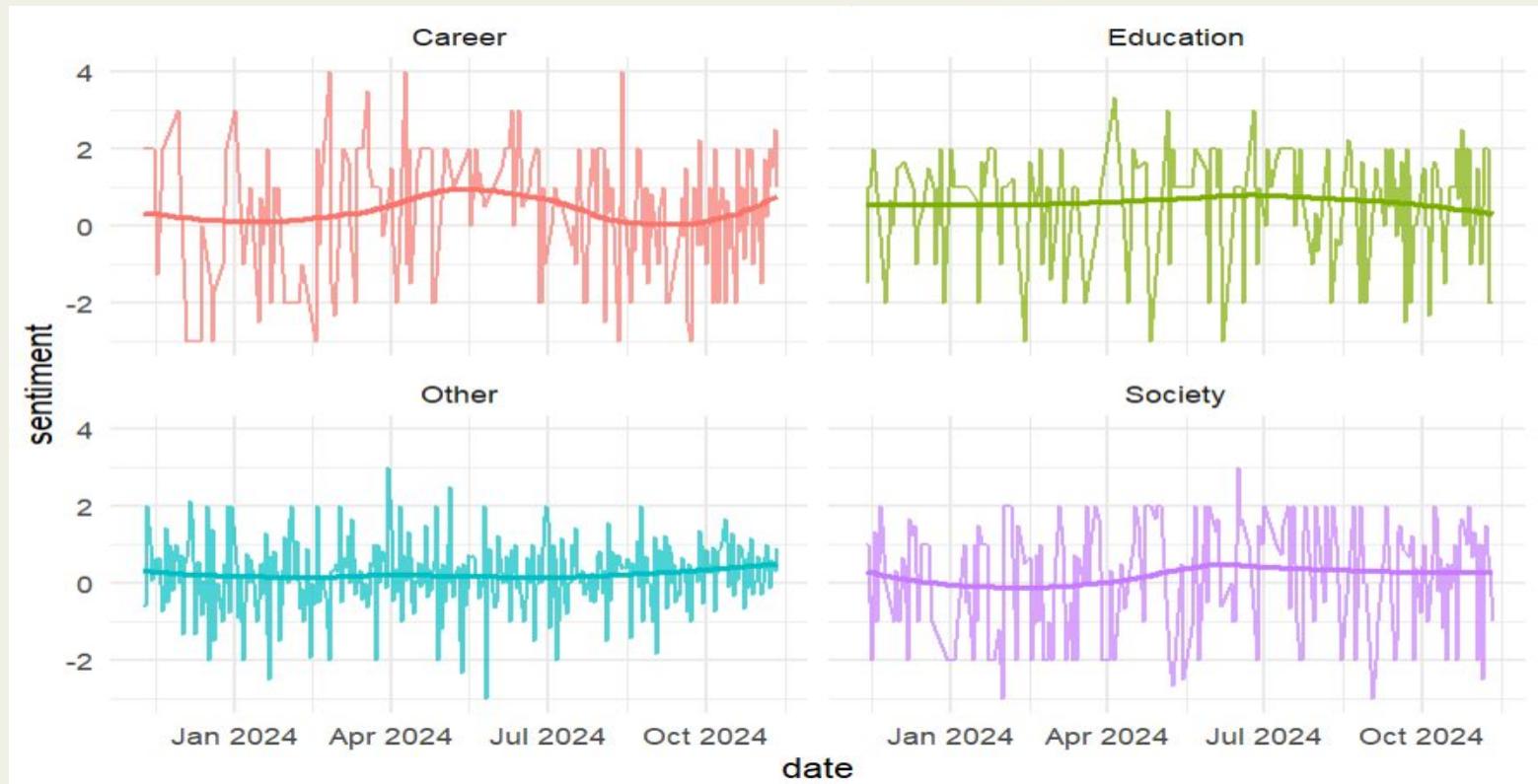
Technologies Used:



Query 1 - Sentiment change over time

- Our first question after looking at the dataset was to determine if there was a **sudden change** in **average sentiment** over time in any of the news categories, or if the sentiment was **fairly uniform**.
- We also wanted to see if the sentiment of news headlines was usually **positive**, **negative** or **neutral** and the valence of those sentiments.
- Technologies Used: ggplot, tidytext, Loess Regression, AFINN lexicon

Visualization #1



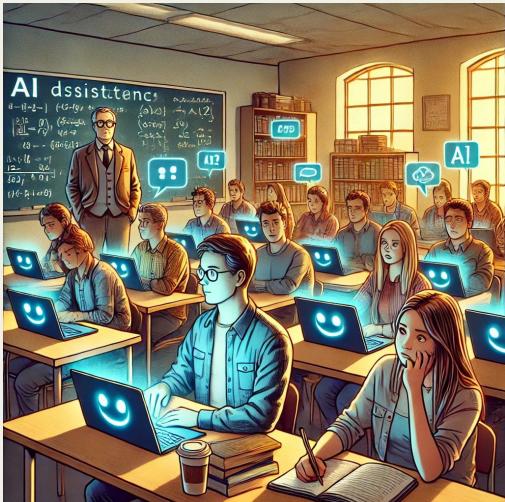
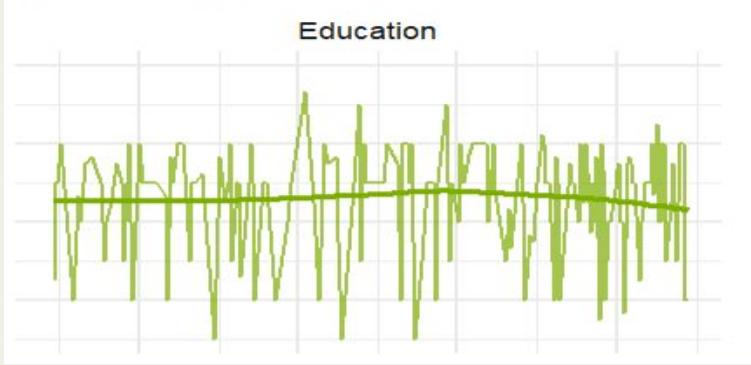
Sentiment Trends in Headlines by Categories

Career



- The highest fluctuations are seen in career, with steeper highs and lows and also a more pronounced curve.
- This indicates that the AI's impact on career is more controversial.
- The Career category shows pronounced highs and lows, indicating that news around AI and jobs can be polarizing—ranging from fears of job displacement to excitement about new AI-driven opportunities.

Education



- The sentiment on education seems to be dropping recently.
- Maybe it is due to concerns of rise of plagiarism and academic dishonesty in college courses and research publications with AI.
- Institutions are increasingly revising policies to address AI usage, emphasizing the need for clear guidelines to maintain academic integrity.

Society



- Overall sentiment has been relatively stable, suggesting that while people are aware of AI's societal impacts, they remain cautiously hopeful about its benefits.
- While the sentiment for society remained stagnant, there is a noticeable rise between late May and Mid-June 2024.
- This may be because of the 2024 AI Seoul Summit held on 21-22 May.
- This event addressed global challenges posed by AI and setting a framework for ethical AI governance, raising confidence in AI's positive impacts on society.

Query 2 - Most common themes in each category

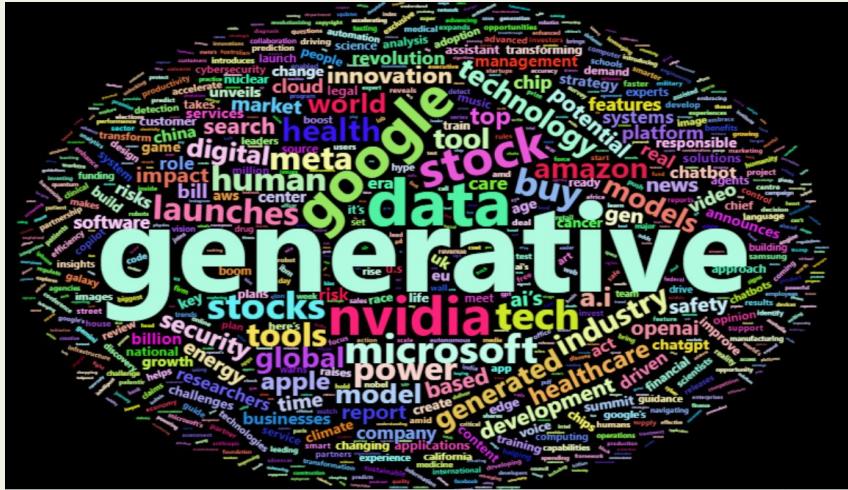
- We now wanted to delve deeper into the main themes in each category.
- Hence, we decided to search for the 5 most common words of each category to deduce that.
- Technologies Used: dplyr, tidytext, ggplot2, stringr, wordcloud2

Visualization #2a

category	word	n	category	word	n
<chr>	<chr>	<int>	<chr>	<chr>	<int>
Career	business	146	Other	generative	443
Career	companies	88	Other	data	234
Career	ceo	83	Other	google	204
Career	startup	70	Other	nvidia	194
Career	generative	64	Other	stock	177
Education	research	161	Society	law	88
Education	learning	128	Society	regulation	68
Education	education	117	Society	public	66
Education	study	89	Society	election	65
Education	machine	74	Society	government	65

Most used words by category

Visualization #2b



- The presence of both established tech giants and emerging startups underscores a dynamic AI ecosystem impacting business innovation and research.
 - Generative AI appears as a recurring theme across multiple categories, underlining its widespread influence.
 - References to law, regulation, and ethical frameworks highlight an increasing public concern over AI governance.

Career

word	n
<chr>	<int>
business	146
companies	88
ceo	83
startup	70
generative	64



- **Business & Entrepreneurship**

“Business,” “companies,” “ceo,” and “startup” underscore AI’s strong linkage to corporate innovation and new venture creation.

- **Generative AI Adoption**

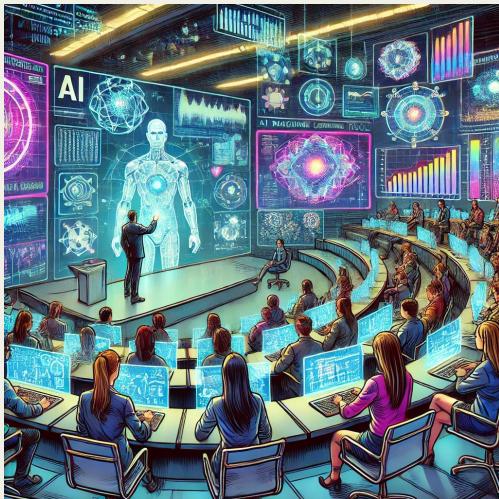
The popularity of “generative” tools (e.g., ChatGPT, Claude) highlights how executive decisions and product strategies are being reshaped.

- **Opportunities & Challenges**

While AI tools create new job paths and industries, they also raise questions about automation and workforce displacement.

Education

generative	64
research	161
learning	128
education	117
study	89
machine	74



- **Research-Driven Discussions**

“Research” stands out as the top mention, reflecting AI’s expanding role in advanced academic and scientific pursuits.

- **Core Focus on Learning**

The words “learning” and “education” confirm continued interest in AI-driven teaching tools and methods.

- **Machine Learning Emphasis**

Frequent references to “machine” highlight the critical role of machine learning in shaping modern pedagogy and study practices.

Society

law	88
regulation	68
public	66
election	65
government	65



- **Legal and Regulatory Concerns**

“Law” and “regulation” dominate, revealing public worries about the misuse of AI and the urgent need for robust governance.
- **Public and Government Involvement**

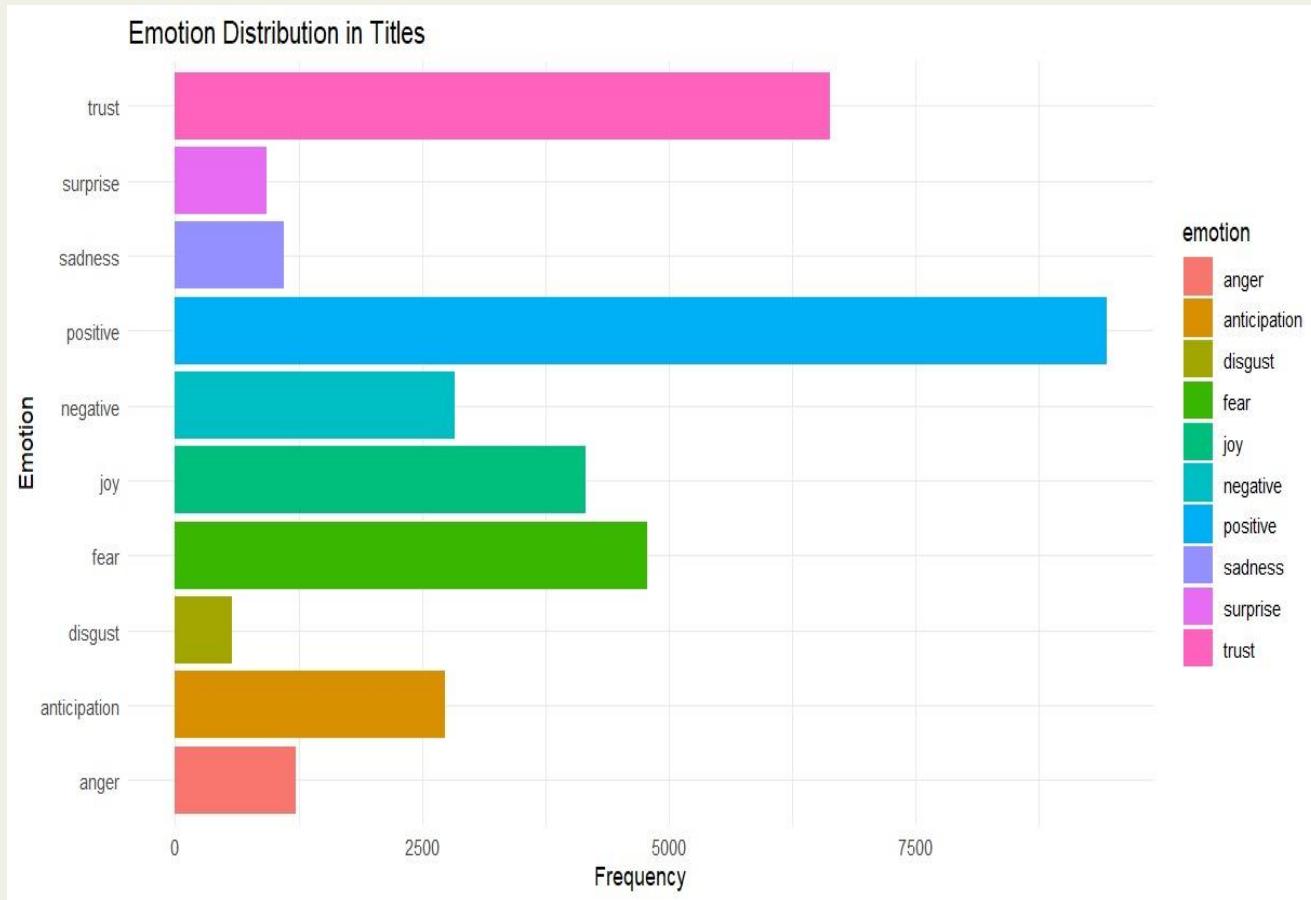
Mentions of “public” and “government” show a broader societal debate on accountability and oversight for AI technologies.
- **Political Implications**

“Election” references suggest AI’s growing role in political campaigns, voter engagement, and policy-making processes.

Query 3 - Emotion distribution in news headlines

- We felt that the sentiment analysis visualization did not give us enough information about the range of emotions expressed in the news headlines.
- For example, emotions like fear and anger that are both categorized as negative are actually quite distinct in the message they convey to us, and could be useful to understand the thoughts of the writer of the headline more accurately.
- Therefore, we decided to analyze these distinct emotions next. Because negative news drives up consumption rates (Robertson CE, Pröllochs N, Schwarzenegger K, Pärnamets P, Van Bavel JJ, Feuerriegel S., May 2023), we are expecting a predominantly negative set of emotions in the following findings.
- Technologies Used: tidyverse, tidytext, lubridate, ggplot2, NRC lexicon

Visualization #3



Findings for Q#3

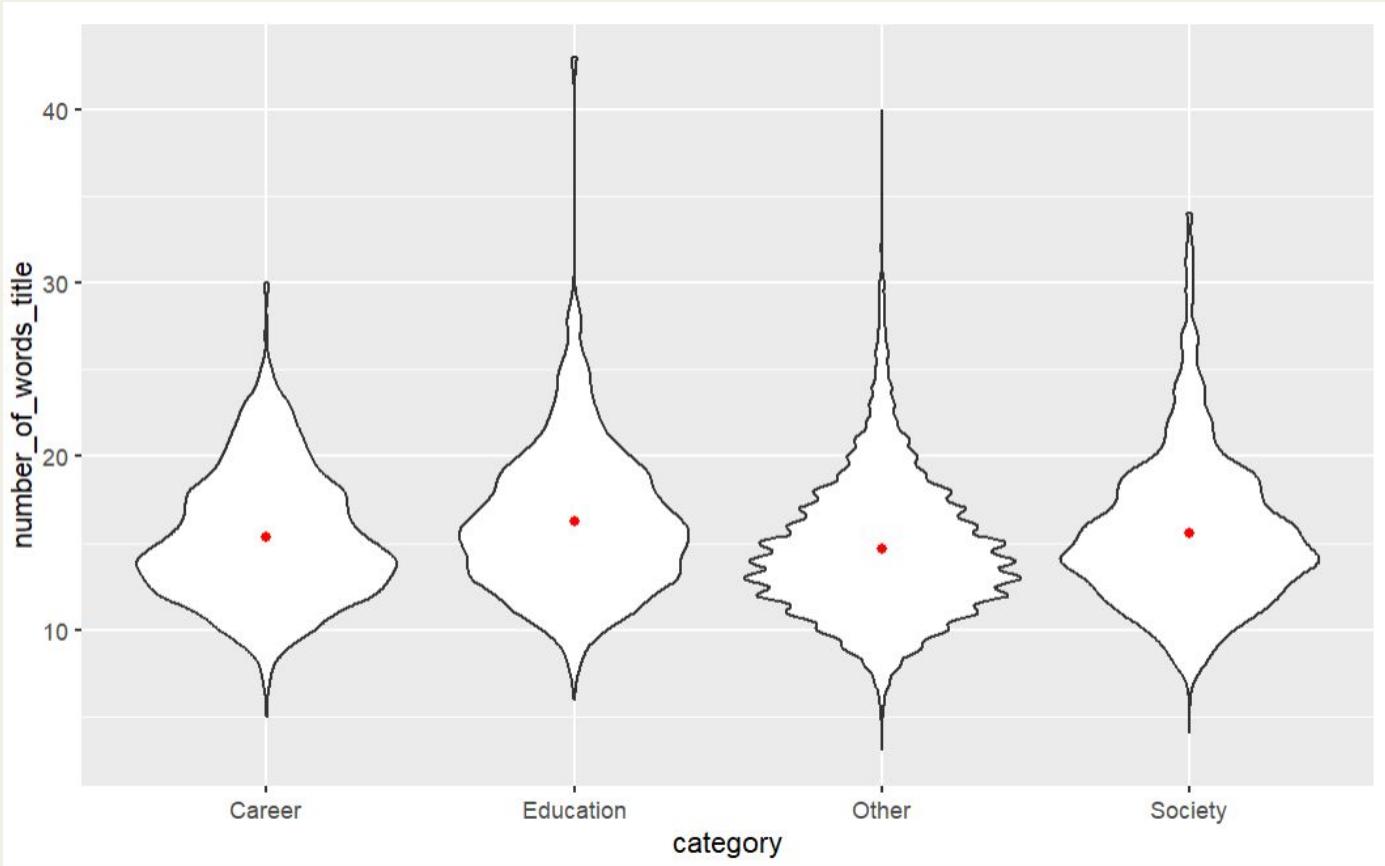
- **Surprisingly Positive Tilt:** Despite earlier expectations of predominantly negative emotions (fear, anger), the headlines are largely positive, challenging the notion of an inherent negative bias in AI news.
 - **High Levels of Trust:** Trust appears in over 60% of the headlines—suggesting that many AI-related stories focus on reliability, innovation benefits, or confidence in new technologies.
 - **Fear Remains Significant:** Nearly half of the headlines convey fear, indicating underlying anxieties around AI's potential risks and its impact on jobs, privacy, and ethics.
 - **Complex Emotional Landscape:** Emotions like anger, disgust, and surprise, though not as frequent, still shape public perception and can signal niche concerns or emerging controversies.



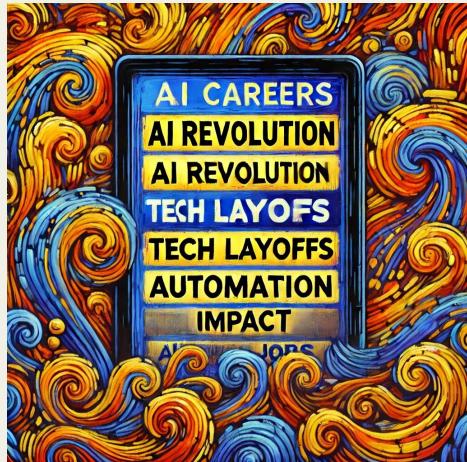
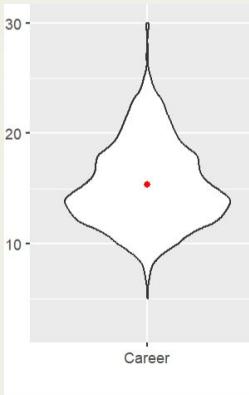
Query 4 - Length of headlines

- Headline length speaks volumes about the engagement tactics used while writing news headlines.
- Shorter headlines may be designed to be more “punchy” and evoke sharper sentiment, being more likely to be “click-bait” because they offer limited context to the reader so as to tempt them into reading the article.
- Whereas longer headlines may be used to provide more context about the news topic while balancing the sentiment by combining both negative and positive tones in headlines to offer a more balanced perspective.
- We are expecting that categories with generally uncertain AI outcomes (like society and jobs) are more likely to have shorter headlines while areas with expectations of mostly positive outcomes (like education) will have longer headlines.
- Technologies Used: ggplot2, dplyr, lubridate

Visualization #4



Career



- **Short, Punchy Headlines**

Headlines in the Career category tend to be concise, possibly tapping into readers' fears of automation and job displacement.

- **Clickbait Potential**

The brevity may be a deliberate strategy to pique curiosity and drive engagement around sensitive career-related topics.

- **Negative Emphasis**

Shorter, sharper headlines often lean on fear-based themes, reflecting public unease about AI's impact on employment.

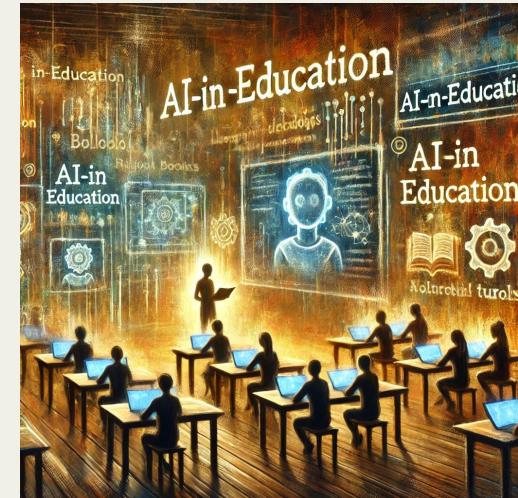
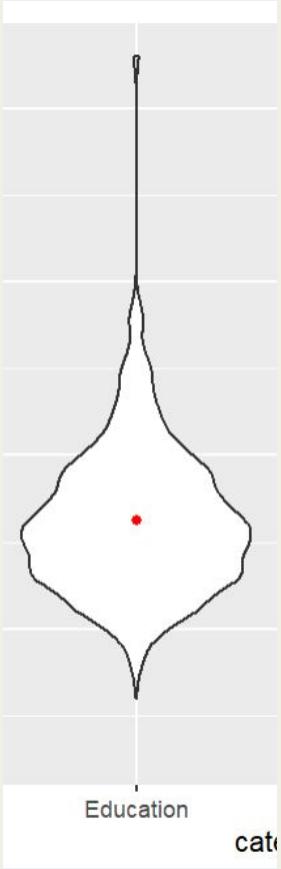
Education

- **Longer Headlines**

Education news tends to provide more words, suggesting a more balanced approach to informing readers about AI's role in learning.

- **Contextual Depth**

By offering additional details, these headlines may aim to clarify both benefits (e.g., innovative teaching tools) and concerns (e.g., plagiarism).



Society

- Concise Headlines, Big Topics

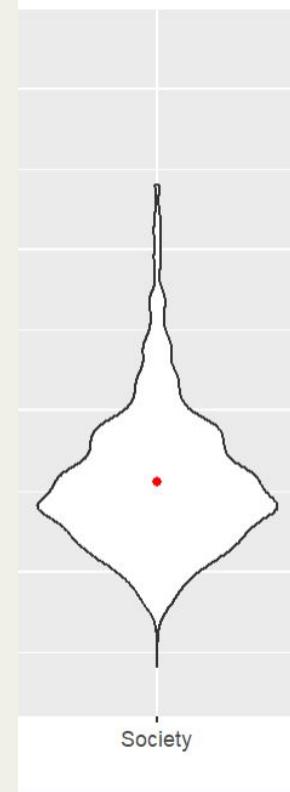
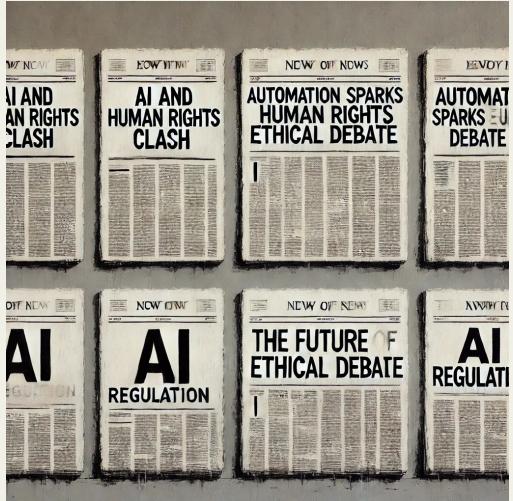
Despite dealing with broad societal implications, many Society headlines remain relatively short and direct.

- Focus on Controversy

Brief headlines may magnify tension around AI ethics, governance, and public policy, drawing in readers who are anxious about societal change.

- Emotional Engagement

Short text often sparks immediate reactions, aligning with earlier findings that fear and negative sentiment drive higher click-through rates.

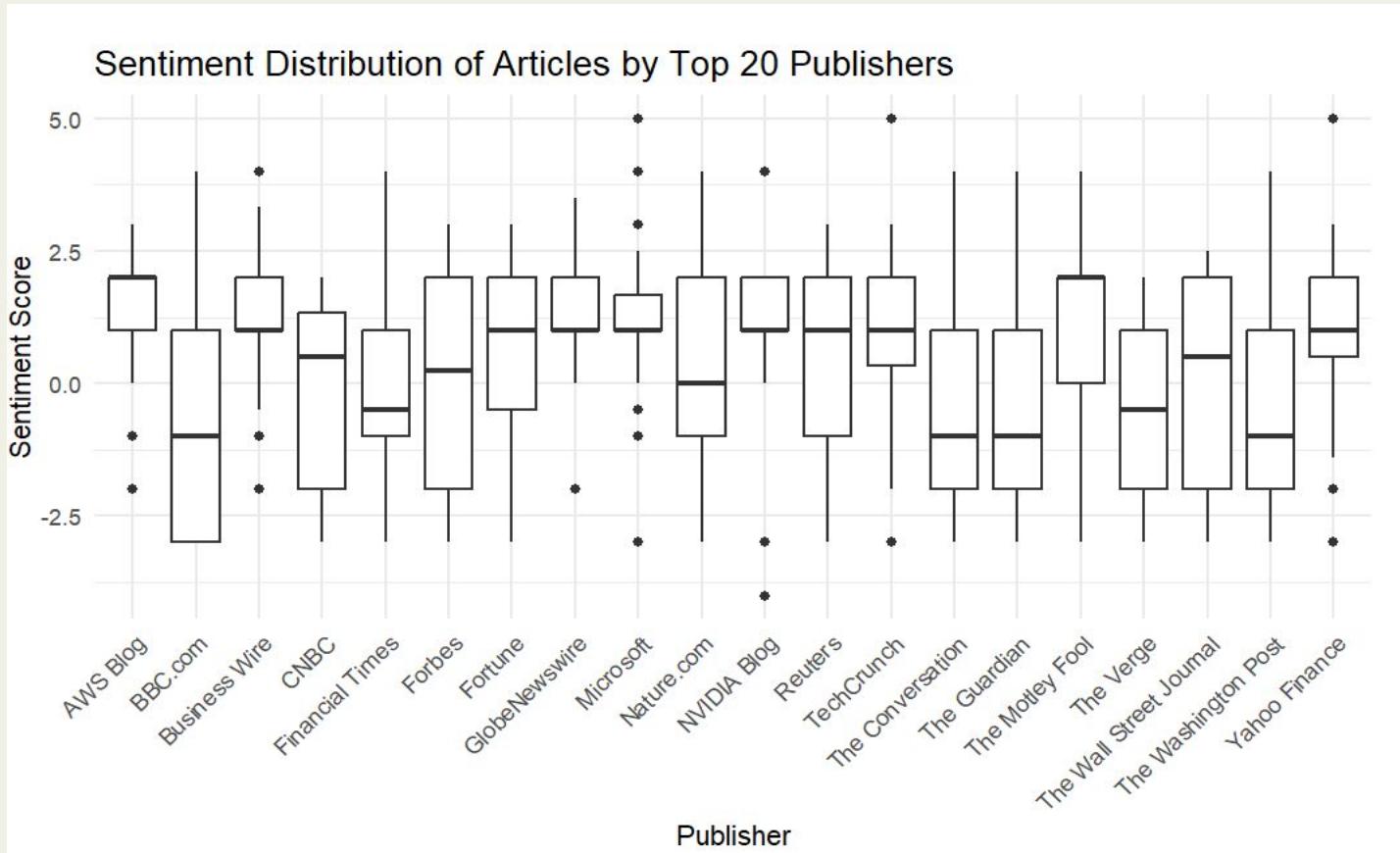


Query 5 - Publisher Bias

- We were wondering if the headlines for news articles are affected by the bias of the publishers or if they are fairly consistent across all their articles
- To investigate this, we can analyze the sentiment of articles published by different sources and compare their distributions in a box plot.
- To test this theory, we calculated the sentiment of articles using the AFINN lexicon, then grouped the results by publisher. By visualizing the sentiment distributions for the top 20 publishers, we can assess whether certain publishers consistently produce more positively or negatively skewed content or if their work averages out to a neutral sentiment.

Technologies Used: dplyr, tidytext, ggplot2, AFINN

Visualization #5



Findings for Q#5

- **Significant bias:** We noticed that there is noticeable bias in how different publishers view and write articles about AI news.
- **Tech Companies are optimistic:** Publishers associated with Big tech companies, such as AWS Blog, Microsoft and NVIDIA Blog, etc. show a highly positive sentiment towards AI, with their lower quartiles consistently positive for AI sentiment.
- **Finance publishers are slightly less positive:** Publishers focused on business and finance news such as Forbes and Yahoo Finance show a positive tilt too, likely due to optimism about AI's positive impact on productivity and the global economy. That said, the sentiment is a little more balanced than tech companies here.



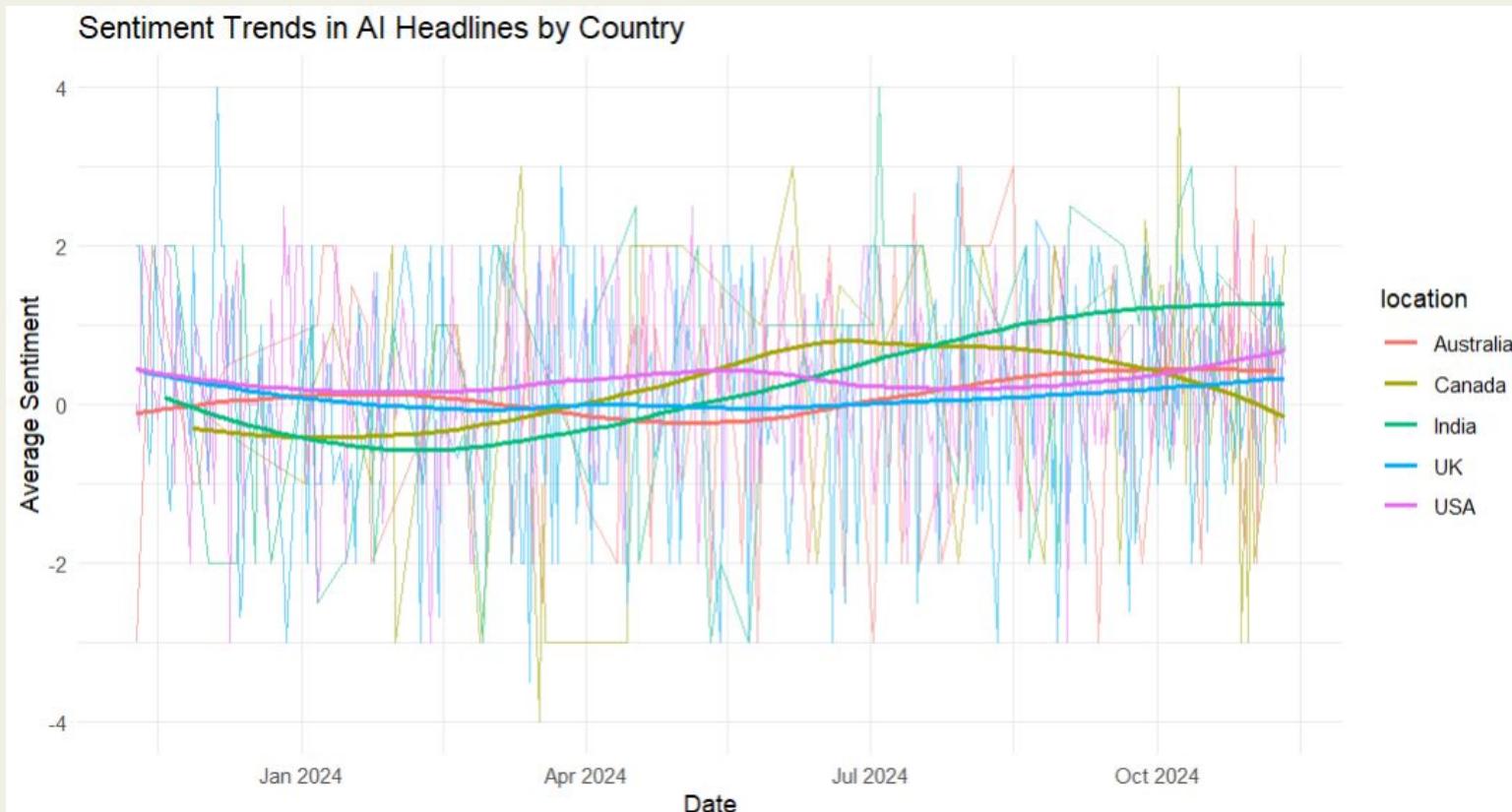
Geographical Analysis

- We used **GPT-3.5** to automatically populate a new country column based on the publisher of each headline. This came at the cost of losing close to **1300** rows of data.
- This allowed us to infer the **country of origin** for most articles giving us geographical context for the dataset.
- With this new column, we unlocked a whole new angle for analysis – similar to how we previously analyzed **categories** of news.

Query 6 - Sentiment change over time by country

- Similar to what we did for the 4 categories, we sorted the top 5 countries with the most news headlines and performed sentiment analysis on them.
- We also wanted to see if the sentiment of news headlines was usually **positive**, **negative** or **neutral** for either of the countries and the valence of those sentiments.
- Technologies Used: ggplot, tidytext, Loess Regression, AFINN lexicon

Visualization #6



Sentiment Trends in Headlines by Countries

Trends

- Sentiment in India seems to be the most rapidly increasing curve, having the lowest point across all the countries and increasing to be the highest.
- Canada's sentiment seems to have peaked around June 2024 and has been dropping since then.
- Sentiments in the other 3 countries have remained relatively stable over time.

Inferences - Canada

- Canadian Government Investment Boosted Sentiment: Canada's announcement of a \$2.4 billion AI investment in April 2024 significantly increased positive sentiment, peaking around June 2024.
- Adoption and Readiness Challenges: Despite initial optimism, low practical AI adoption (only 6.1% of businesses) and decreased readiness due to talent shortages and infrastructure gaps likely contributed to declining sentiment.
- Regulatory and Competitive Pressures: Uncertainty from the failure of the AI regulatory bill (AIDA) and increased global competition, notably from cost-effective AI solutions abroad, further dampened confidence in Canada's AI sector.

Inferences - India

- Policy Shift and Sentiment Dip: In Feb-March 2024, India's AI sentiment saw a notable decline following government's advisory mandating approval for deploying generative AI models. This move faced criticism from industry leaders for potentially hindering innovation. It was subsequently revised to emphasize self-regulation and industry collaboration.
- IndiaAI Mission's Positive Impact: The launch of the ₹10,372 crore (~\$1.25 billion) IndiaAI Mission in March 2024 marked a significant investment in AI research, infrastructure, and startups. This initiative, focusing on enhancing compute capacity, datasets, and innovation hubs, played a crucial role in restoring confidence and driving positive sentiment in India's AI landscape.

Query 7 - Most common themes in each country

- Similar to what we did for the 4 categories, we wanted to find the most common words by country.
- Again, we chose to go with the 5 most common words, of the top 5 countries.
- Technologies Used: dplyr, tidytext, ggplot2, stringr, wordcloud2

Visualization #7

	location	word	n
1	Australia	data	23
2	Australia	australia	21
3	Australia	australian	15
4	Australia	generative	11
5	Australia	responsible	11

21	USA	generative	289
22	USA	stock	174
23	USA	nvidia	170
24	USA	data	141
25	USA	google	128

16	UK	uk	59
17	UK	generative	54
18	UK	data	45
19	UK	google	42
20	UK	tech	37

11	India	india	18
12	India	google	14
13	India	learning	10
14	India	data	9
15	India	role	9

6	Canada	data	11
7	Canada	stocks	11
8	Canada	canadian	10
9	Canada	canada	9
10	Canada	generative	9

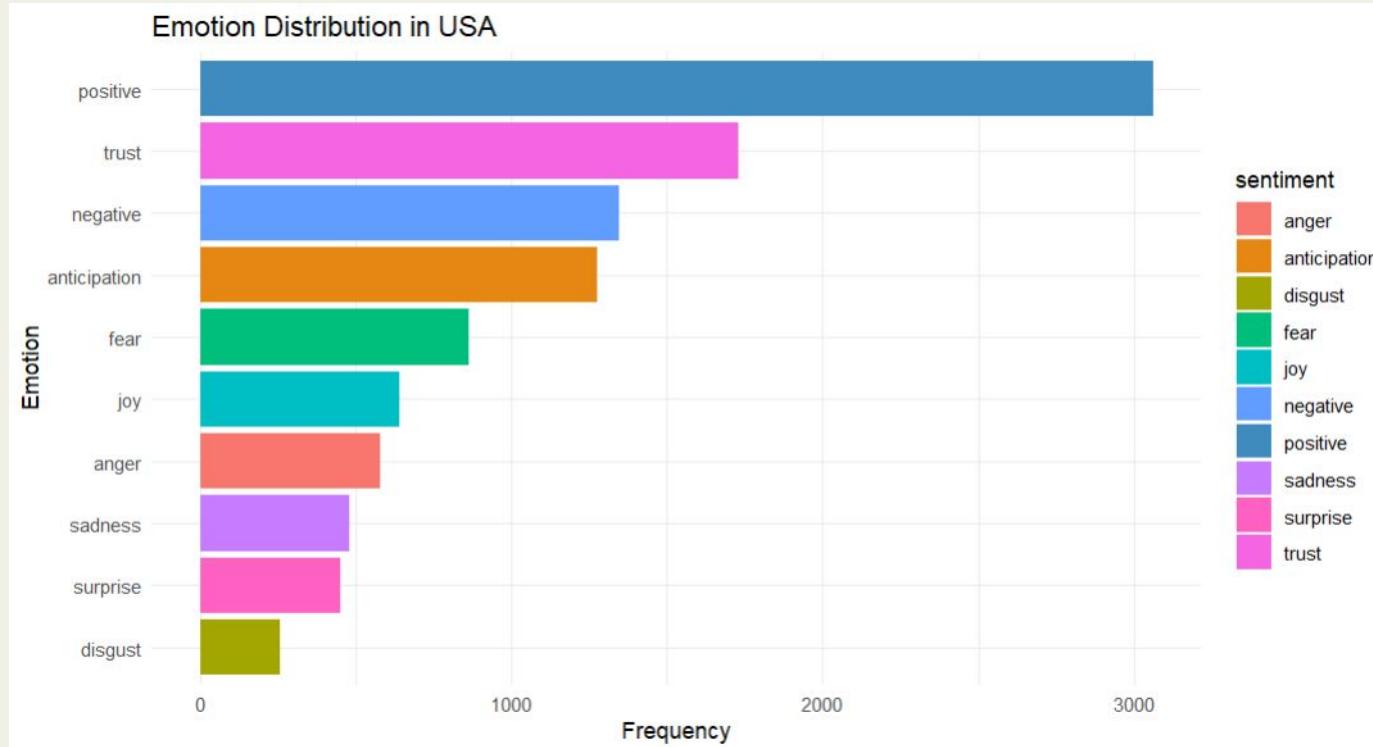
Trends

- Apart from the common words like generative, data, etc. All the countries' news headlines commonly featured the names of those countries, with the exception of USA.
- “Google” was commonly featured in Indian, British and American headlines.
- USA and Canada's news headlines commonly featured stocks as a point of interest.

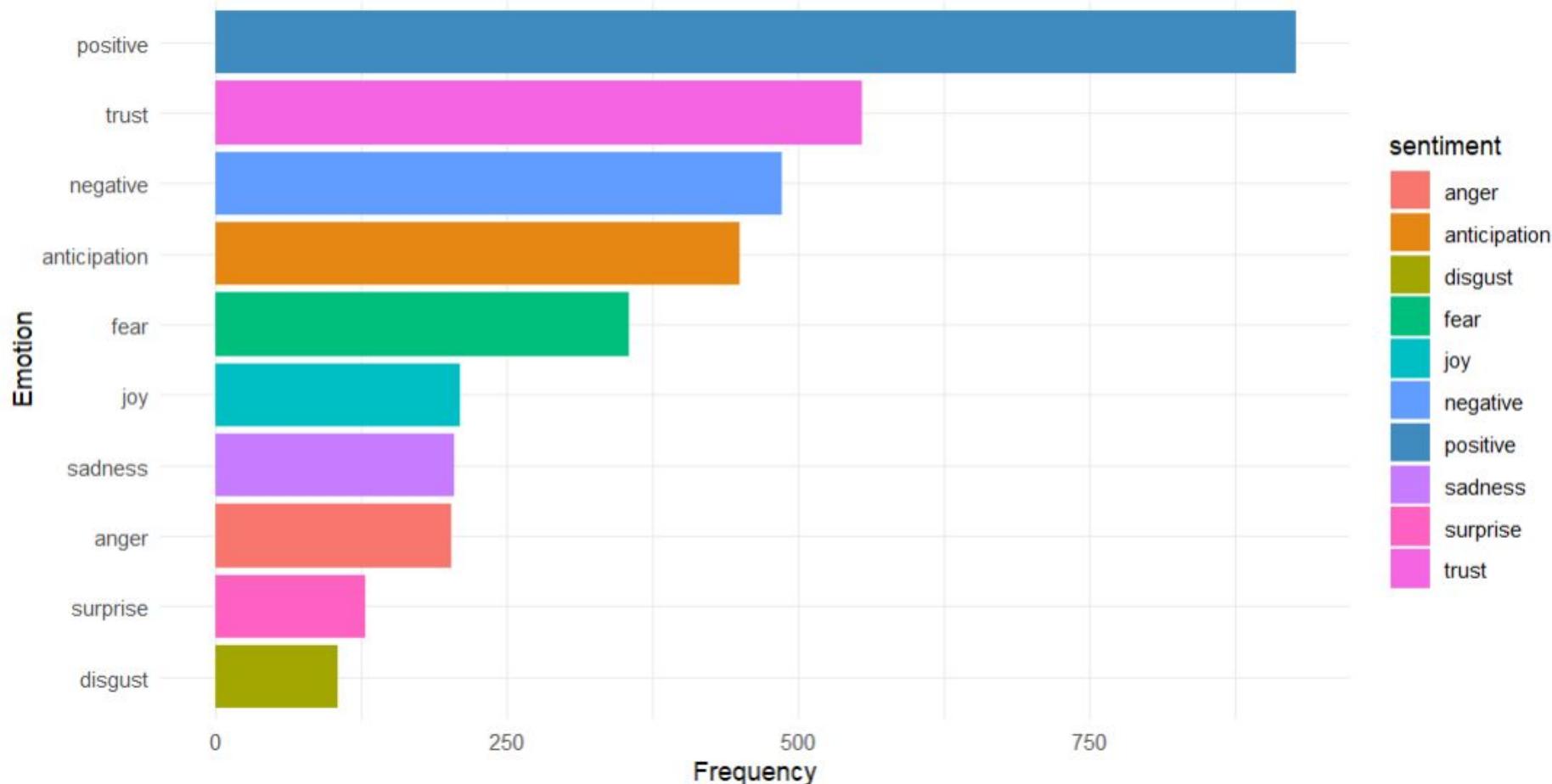
Inferences

- Most countries' AI news headlines emphasize national identity and local achievements, with frequent mentions of their own names, while U.S. headlines focus on industry giants like Nvidia and Google, and market dynamics like "stock," without referencing "America" or "USA."
- The U.S.'s unique narrative reflects its status as a global AI leader, where corporate influence and financial markets overshadow national context, unlike countries like India and the UK, which tie AI developments to national strategies and local tech players.

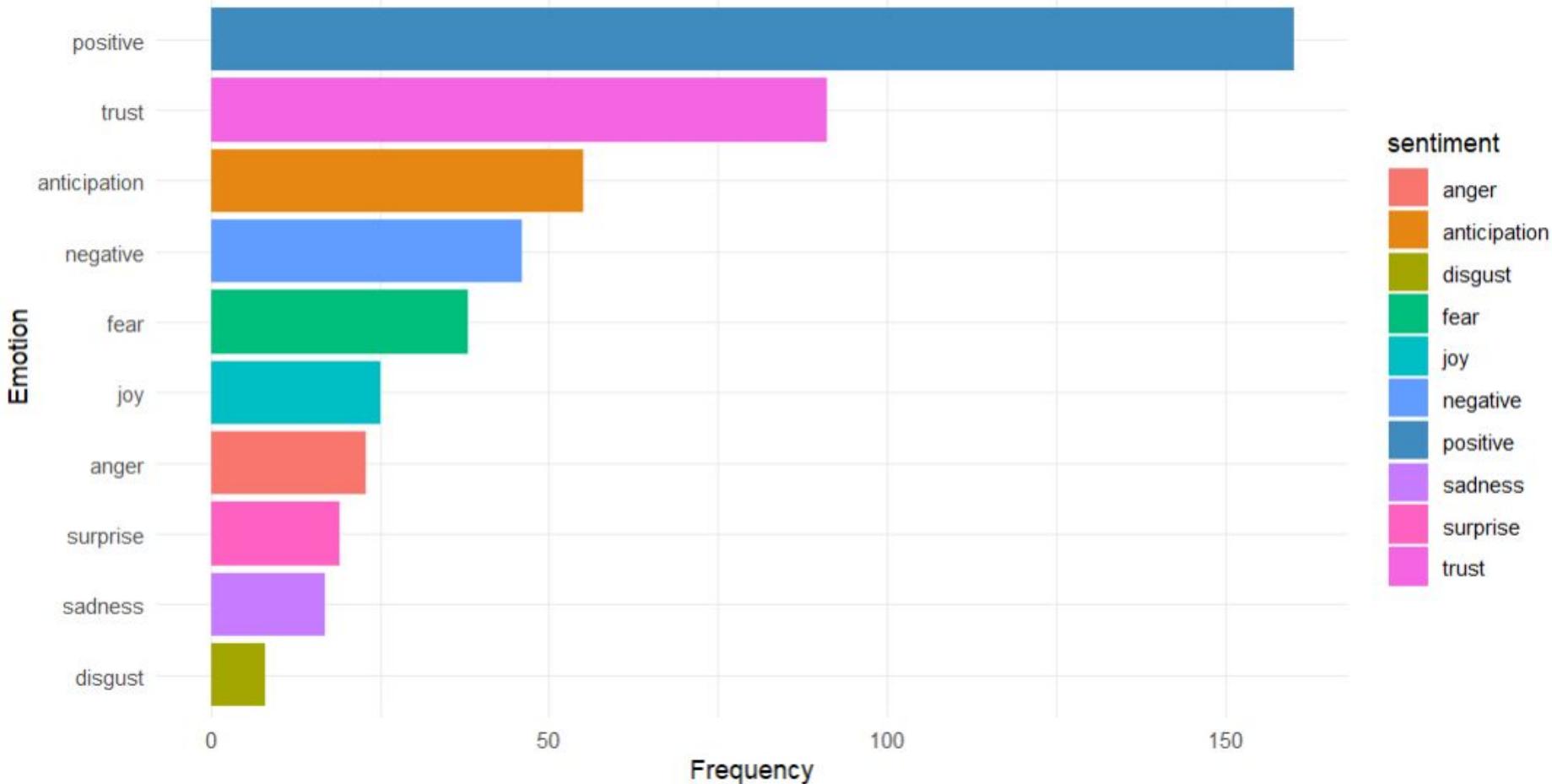
Query 8 - Emotion distribution in news headlines across the top 5 countries.



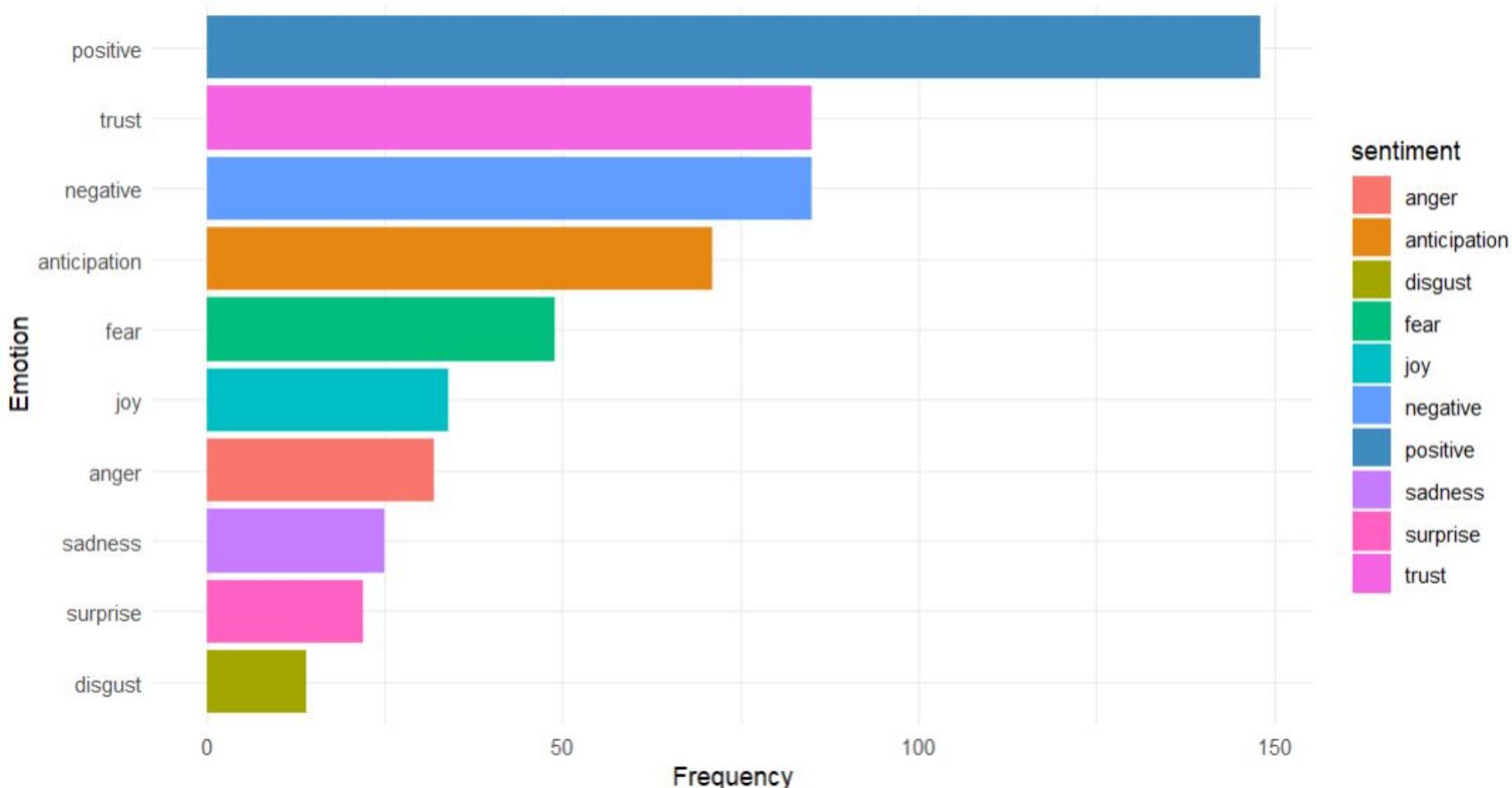
Emotion Distribution in UK



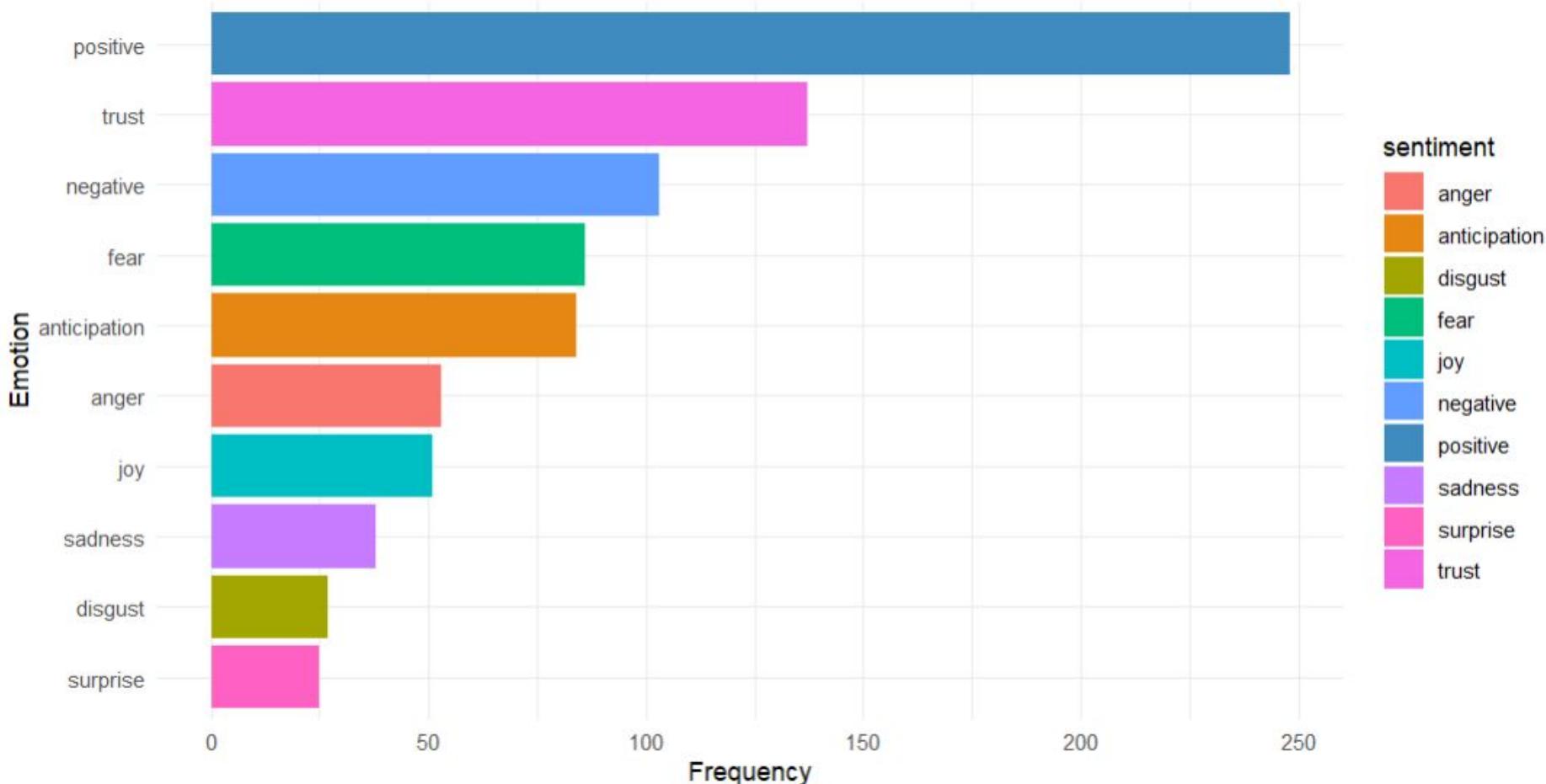
Emotion Distribution in India



Emotion Distribution in Canada



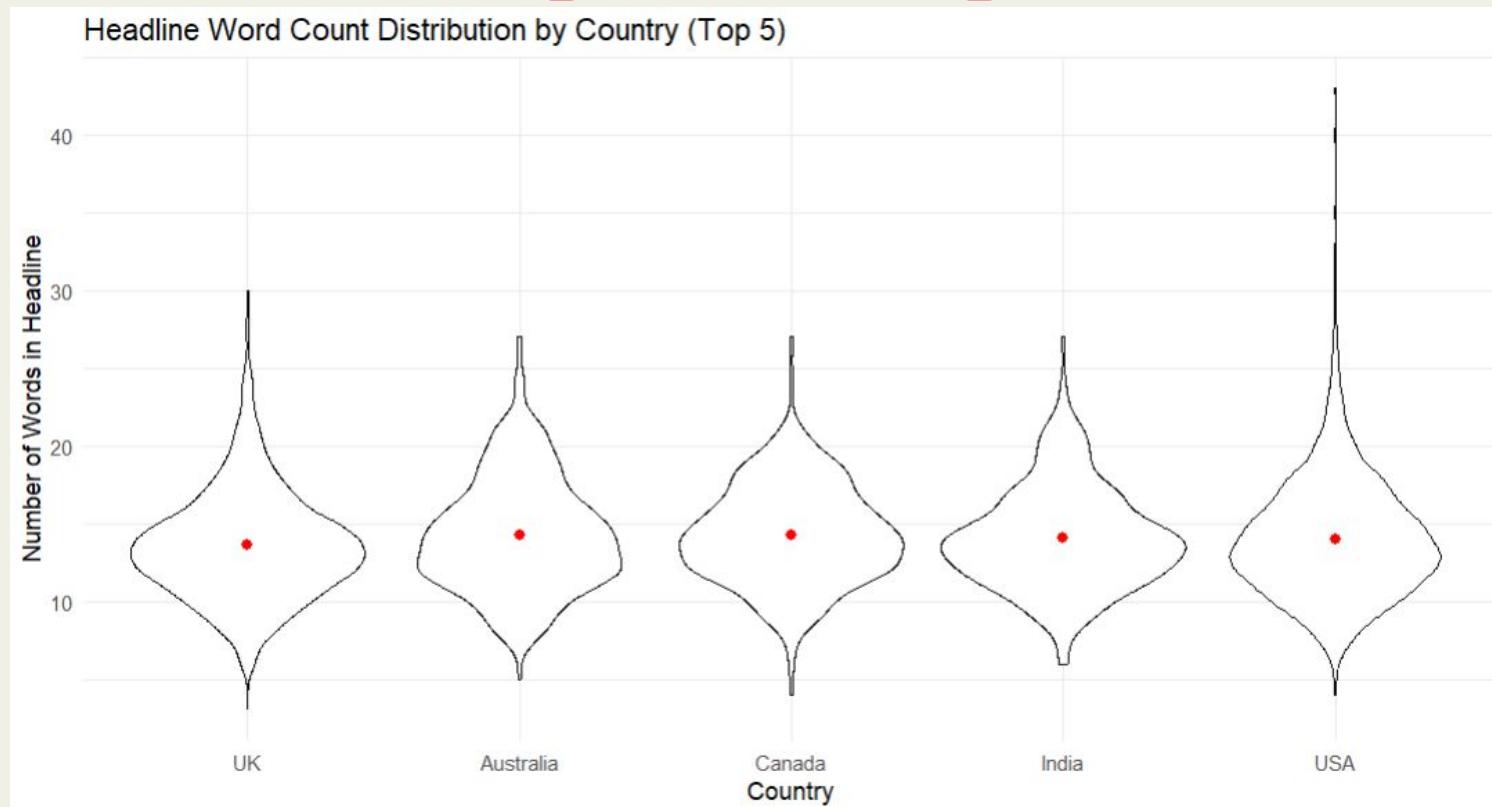
Emotion Distribution in Australia



Trends

- Similar to the findings in the earlier analysis with the categories, the news headlines mostly have the positive emotion in them, followed by trust, negative, anticipation, etc. And this is mostly same for every country.
- Corroborating our findings in the sentiment analysis, India seems to have a somewhat larger amount of positive emotions as compared to the other countries due to the aforementioned reasons.
- Similarly, Canada has a larger percentage of anticipation and negative sentiment in news headlines as compared to the other countries which aligns with our earlier analysis and reasoning.

Query 9 - Length of headlines by country



Trends

- All countries except for USA share a similar distribution of headline word count(could be exaggerated because of the higher volume of data from USA), with USA having the most variance and the most outliers.
- Indian news headlines have the least variance in word count, shortly followed by the UK. This could be due to language and media-based cultural conventions in those countries compared to the rest of the countries.

Conclusion

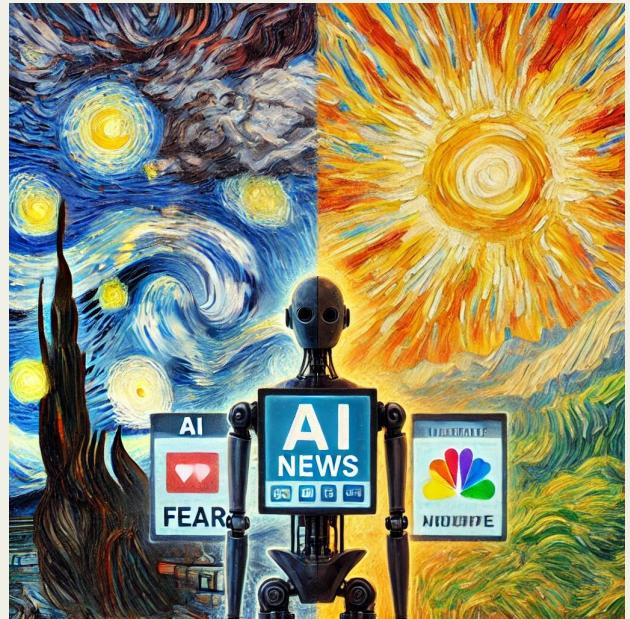
- **Positive Tilt Overall:** Contrary to expectations of a negative bias, AI headlines show a predominantly positive sentiment, with trust and fear being the most common emotions.
- **Generative AI's Widespread Impact:** Across categories, “generative” emerges frequently, highlighting its transformative influence in business, research, and public perception.
- **Regulation & Societal Concerns:** Topics like law, regulation, and ethics underscore ongoing debates around responsible AI governance and public unease about potential misuse.
- **Headline Length Strategies:** Shorter, more “clickbait” headlines dominate in categories with higher uncertainty (Career, Society), while Education headlines tend to be longer, suggesting a more balanced, informative approach.

- **Publisher Bias:** Tech-focused sources (AWS, Microsoft, NVIDIA) present consistently optimistic AI views, while finance and general news outlets adopt a balanced stance. Understanding these biases ensures equitable AI governance.
- **Regional Sentiment:** India's sentiment rebounded strongly after significant AI investments like the IndiaAI Mission; Canada's sentiment fell amid regulatory challenges. Tailored local policies are crucial.
- **Public Emotions:** Although overall sentiment is positive, persistent fear and skepticism highlight public concerns. Ongoing sentiment monitoring can help inform proactive media literacy and balanced reporting efforts.

References

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature human behaviour*, 7(5), 812–822.
<https://doi.org/10.1038/s41562-023-01538-4>

Garvey, C.S., & Maskal, C. (2020). Sentiment Analysis of the News Media on Artificial Intelligence Does Not Support Claims of Negative Bias Against Artificial Intelligence. *Omics : a journal of integrative biology*.



Thank You