TRUMP IN THE HEADLINES: ANALYSIS OF POLITICAL DISCOURSE IN THE GUARDIAN

KOFI BARTON-BYFIELD
24375742

Contents

Abstract

This paper examines the evolution of media sentiment toward Donald Trump during his campaign for a second non-consecutive term, and its relationship to public opinion and journalistic focus. Using natural language processing, the study analyses headlines from The Guardian between November 2022 and April 2025, focusing on sentiment analysis and topic modelling via Latent Dirichlet Allocation (LDA).

The findings suggest that media sentiment, rather than leading public opinion, tends to reflect it with a slight delay. While media coverage often emphasises Trump's character and controversies, policy-related coverage increases as elections approach. Contrary to expectations, media sentiment does not become increasingly polarised during key campaign milestones; instead, it may become more neutral, reflecting a shift towards policy-focused reporting.

The study concludes that the interplay between media, public opinion and political narrative is complex, with media acting more as a mirror than a driver of public sentiment.

1 Introduction

The media's portrayal of political figures significantly shapes public perception and influences electoral dynamics. Among contemporary politicians, Donald J Trump's controversial policies and impulsive communication style have consistently attracted intense media scrutiny. His announcement of the 2024 US presidential campaign in November 2022 presents a rare and revealing case study. Unlike typical presidential candidates, Trump is the only modern US president seeking a second non-consecutive term. A scenario not seen since Grover Cleveland in the 19th century. This means he re-enters the political arena not as a newcomer or sitting president, but as a highly polarising figure with a well established media image and a legacy. As such, the 2024 race provides a unique opportunity to examine how prior presidential performance shapes contemporary media narratives. Along with how those narratives may, in turn, influence a campaign led by a known and controversial character.

1.1 The Literature

People, particularly in the United States, often vote along party lines or based on specific partisan issues they associate with their chosen political affiliation. This behaviour is deeply influenced by the candidate's ability to project a confident, consistent image, with voters gravitating towards those they perceive as "authentic" or "true to themselves." Such qualities are especially compelling when contrasted with the typical perception of mainstream politicians as untrustworthy, inconsistent, or self-serving [Breitenstein et al.2025]. Authenticity, therefore, becomes a key factor in determining how voters perceive and relate to political figures, influencing their likelihood to support candidates who seem to embody their values and beliefs. This dynamic is particularly significant in an era where the public's trust in politicians is at a historical low, making the image of a genuine, unfiltered candidate increasingly appealing [Liu2024].
While the influence of social media echo chambers on voting behaviour has been widely documented [Frimpong et al.2022], traditional media outlets such as newspapers still play a significant role in shaping political opinions, particularly among older demographics and more institutional audiences [Dada2023]. Unlike the fast-paced, often polarising nature of social media, traditional media provides a more curated narrative that can deeply influence individuals who still rely on print journalism for their political information. These audiences tend to be more susceptible to the framing and tone established by respected institutions, which can subtly sway political opinions over time.
It is well known that media coverage increases in volume during the lead-up to elections, with both the quantity and intensity of coverage rising dramatically as candidates compete for public attention [Hopmann et al.2012]. Alongside this surge in media presence, the content itself often shifts subtly in line with the electoral calendar [Takens et al.2013]. In the United States, for example, the framing and substance of media coverage frequently reflect the political bias of the news outlet, with certain publications favouring particular candidates or policy positions [Berning2023]. Although this bias may not always be overt, it plays a crucial role in shaping both the tone of journalistic reporting and the broader public narrative around electoral campaigns. This subtle influence can significantly affect voter perceptions, contributing to a more polarised electorate, where voters' choices are often guided not just by policy, but by how candidates and their platforms are portrayed through the lens of media coverage.

1.2 Research Question

From the literature the following research question was formed:

> **"How does media sentiment towards Trump evolve throughout his campaign for a second non-consecutive term, and what does this reveal about journalistic focus and public opinion?"**

This question explores how media tone shifted during major points in the campaign and whether those shifts mirrored or influenced public opinion. It also looks at whether coverage leaned more towards Trump's character or his policies; and what that says about the way the media shaped the story around him.

1.2.1 Motivation & Hypotheses

To address this research question, three key hypotheses have been formed, each designed to explore a distinct aspect of the relationship between media coverage and public or journalistic response to Trump's campaign.

H1: Shifts in media sentiment influence changes in public opinion.

H2: Media coverage of Trump is predominantly character focused rather than policy focused.

H3: Media sentiment towards Trump becomes increasingly polarised during key campaign milestones.

These hypotheses explore different aspects of how media coverage may have influenced or reflected public and journalistic attitudes towards Trump's campaign. **H1** suggests that changes in media sentiment, whether positive or negative, could have led to shifts in public opinion, implying that the media played a key role in shaping voter perceptions. **H2** focuses on the type of media coverage, arguing that the press predominantly emphasised Trump's character and personality over his actual policies, which may have contributed to the public's view of him more as a figure than a politician. Finally, **H3** looks at the intensity of media sentiment over time, proposing that coverage became more polarised and extreme in its praise or condemnation during significant campaign events. This shift in coverage reflects a growing ideological divide in how Trump was portrayed.

2 Methodology

2.1 The Data Sources

The data for this paper was sourced from The Guardian, a reputable, centre left news publication with an accessible API. This API enabled the creation of a dataset of articles relevant to Donald Trump's political activity. The collection period spans from `2022-11-15` to `2025-04-13`, commencing with the date of Trump's official announcement of his candidacy for the 2024 presidential election. The API query was constructed with the following parameters to ensure the retrieved articles were pertinent to the research focus:

```
params = {
    "from-date": ‘2022-11-15’,
    "to-date": ‘2025-04-13’,
    "api-key": API_KEY,
    "page-size": 200,
    "page": 1,
    "q": "Trump",
    "query-fields": "headline",
    "show-fields": "body, headline"
}
```

This query specifically targeted articles containing “Trump” in their headlines. A total of 5,288 articles were extracted, providing a substantial dataset for a robust analysis. The restriction to headlines ensures an accurate framing of Trump from the Guardian.

To operationalise the sentiment the headlines will be tokenised and stripped of stop words before being grouped into week commencing. This provides 127 weeks worth of opinionated news commentary on Trump. This data set will be the key data for the topic analysis.

Figure 1 illustrates the weekly distribution of these Trump headline articles. As the graph depicts, there is significant variation in the volume of articles. Notably, the number of articles generally increases as the 2024 election approaches, with a pronounced spike around the time of Trump's inauguration and hush money scandal. This

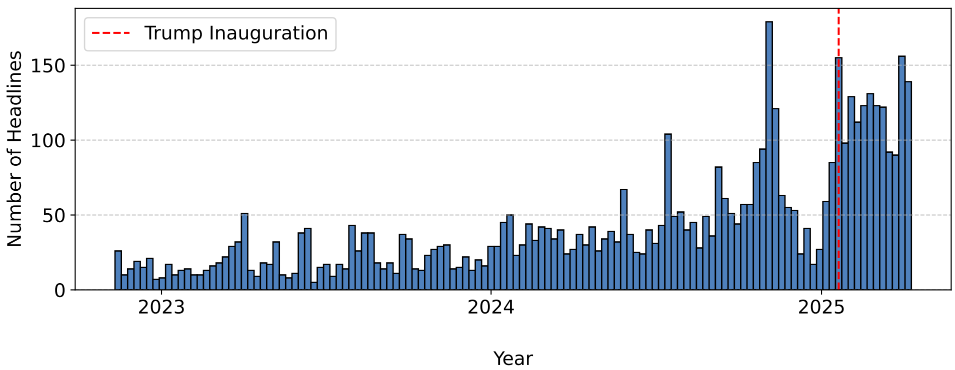

Figure 1: Week Occurrence of Trump Headline Articles

trend suggests that media coverage intensity is at least partially influenced by key events in the election cycle. The data also exhibits what appears to be regular volatility, perhaps tied to other news cycles. Overall, the dataset's size and distribution offer a solid foundation for exploring how The Guardian portrays Donald Trump.

2.2 Theoretical Concepts

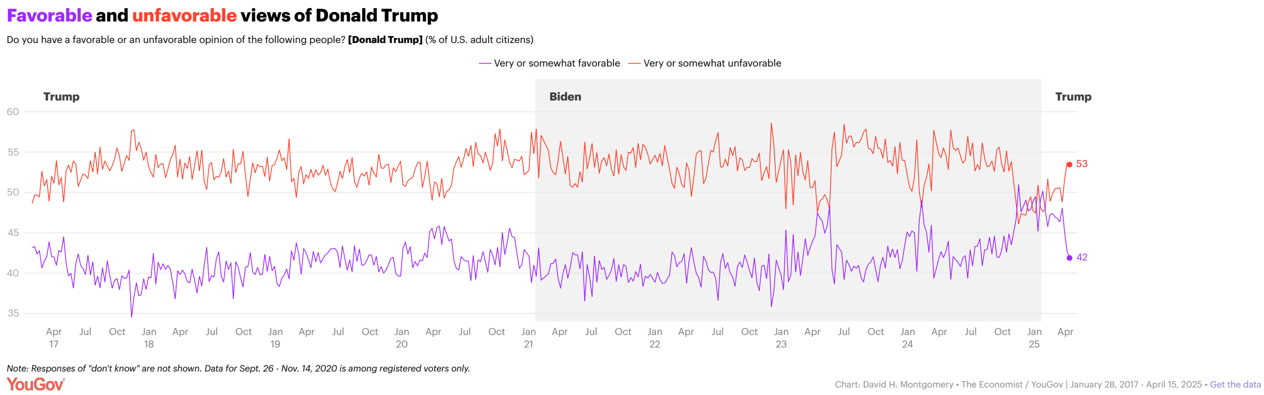

Figure 2: YouGov Opinion Poll

To conceptualise **H1** we need a way to measure the influence of the Guardian articles. To provide this comparative perspective on the media portrayals of Donald Trump, this study will incorporate data from public opinion polls. Specifically, [YouGov2025, April 9] opinion poll data, which tracks Trump's popularity ratings over time, will be used as shown in Figure 2. By comparing the trends in media sentiment, derived from the Guardian newspaper articles, with the fluctuations in public opinion we are able to gain a better understanding of the interplay between the two.

2.3 Strategy & Methods

Latent Dirichlet Allocation (LDA) extracts the top topics mentioned in the text by uncovering patterns in the words that frequently occur together across a collection of documents. It assumes that each document is a mixture of various topics and each topic is represented by a distribution of words. By applying this model, we can identify the underlying themes in the media coverage of Trump's campaign.

LDA operates on the premise that each document in a corpus is a mixture of a finite number of topics and each word in a document is attributable to one of these topics. The goal of LDA is to reverse-engineer the topics that best explain the observed words in the documents, based on the assumption that words occurring together often have a common underlying theme. Through this process, LDA can provide insights into the key issues, opinions and sentiments that are being discussed along with how these evolve over time.

By examining these topics we can not only gain a better understanding of the media coverage of trump **H2**, but also measure the change in sentiment over time as the election and its results approach **H3**.

3 Discussion

3.1 Sentiment Analysis

The net sentiment for each Guardian article was calculated by subtracting the negative sentiment score from the positive sentiment score. These scores were calculated for each article using the `NLTK SentimentIntensityAnalyzer`. To mitigate the impact of short-term volatility and emphasise longer-term trends in media sentiment, a rolling average was applied to the weekly net sentiment values. This smoothing technique reduces noise, allowing for a clearer observation of the underlying sentiment trajectory in relation to public opinion trends.

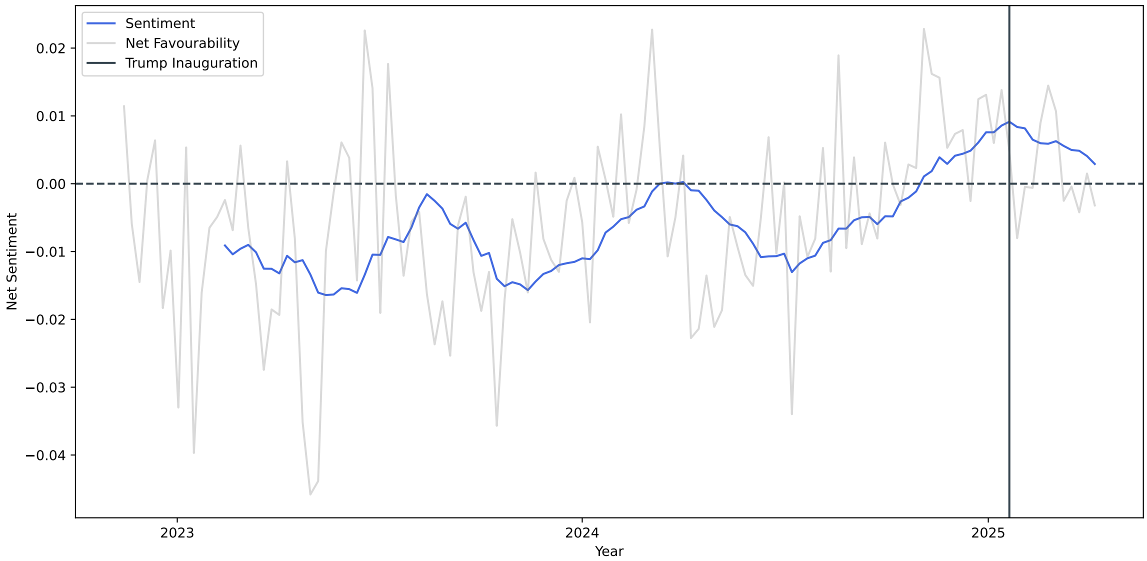

Figure 3: Weekly Sentiment on Trump From the Guardian

Here, we clearly see three distinct peaks, each corresponding to periods where Trump's popularity surged. These coincide with events such as his Republican nomination, key televised debates and the election itself.
Figure 4 below shows opinion poll ratings from [YouGov2025, April 9], using the same 'Net Favourability' metric described earlier.

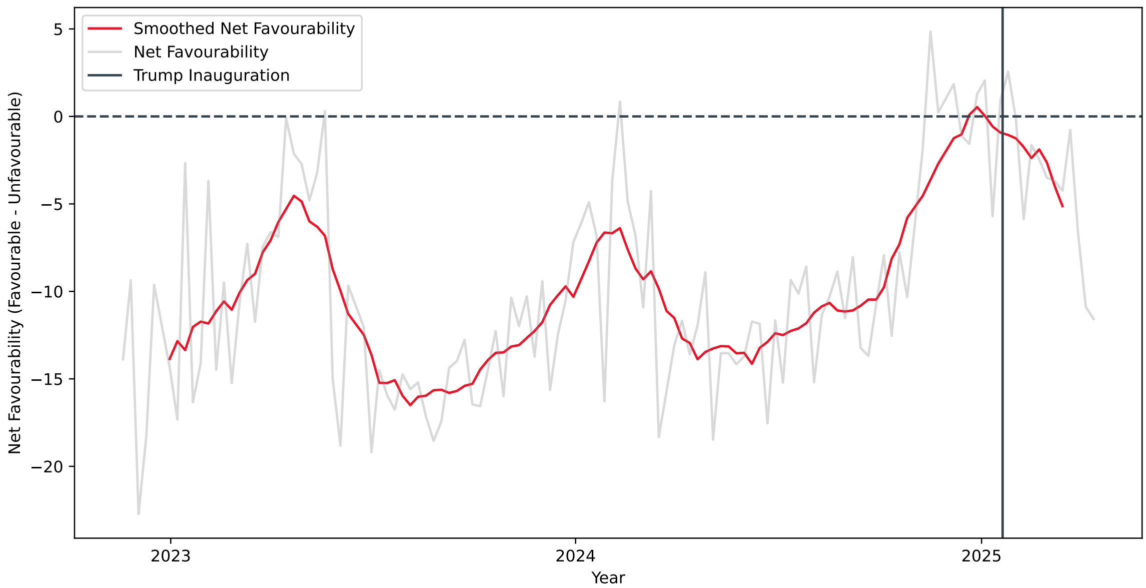

Figure 4: YouGov Opinion Poll

The inclusion of a vertical line on both graphs, denoting the date of Trump's inauguration, provides a key point in time for assessing the relationship between media sentiment and public opinion before and after this significant political event. Notably, the two time series exhibit a broadly similar shape, characterised by three gradually increasing peaks. This alignment in overall trend provides confidence in the efficacy of the sentiment assignment process using the `NLTK SentimentIntensityAnalyzer`. Furthermore, the alignment offers preliminary evidence suggesting that Guardian articles may reflect a lagged public perception of Trump, with shifts in media sentiment potentially following changes in public opinion.

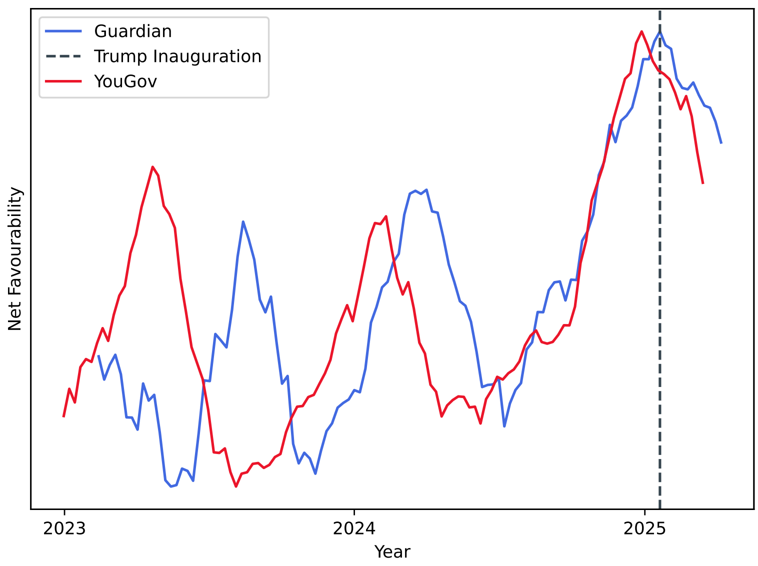

Figure 5: Guardian Sentiment with YouGov Polling

As illustrated in Figure 5, when rescaling the favourability axis, the Guardian sentiment trend appears to follow the YouGov polling data, exhibiting a lagged relationship. The Guardian's sentiment trend becomes more aligned with the YouGov polling as the 2024 presidential announcement approaches. This suggests that media sentiment may become more reflective of public opinion as the election cycle intensifies as in **H3**. This increasing alignment warrants further investigation into the evolving dynamics between media reporting and public opinion as an election nears. It is possible that during periods of heightened political attention, such as the run-up to a campaign announcement, the media focuses more on reflecting and amplifying existing public sentiments. This could be due to a greater reliance on polling data, increased efforts to capture the public mood or a shift in editorial strategy. To further evaluate the key cause of this change the topics discussed must be addressed.

3.2 Latent Dirichlet Allocation (LDA)

To model the topics discussed in The Guardian's articles about Donald Trump, I performed Latent Dirichlet Allocation (LDA) on the cleaned tokens from the articles. As shown in Figure 1, after tuning, the LDA model identified four distinct topics in the headlines.

Topic ID	Keywords	Contextualised
0	Haley, Georgia, Carroll, Willis, county	Legal Proceedings
1	Daniels, Michael, Merchan, Pecker, Stormy	Stormy Daniels
2	tariffs, government, administration, Ukraine, war	Global Trade and Conflict
3	Kamala, debate, today, rally, government	Campaigning

Table 1: Topics Identified in the Analysis

Topic 0: This topic features Haley, which refers to Nikki Haley, the American politician and diplomat who was running directly against Trump for the Republican nomination in the 2024 election. Additionally, Haley was a vocal supporter of E. Jean Carroll, who was awarded $5 million after winning her sexual assault case against Donald Trump. The presence of “Willis”, “Georgia,” and “County” references the case of *The State of Georgia v. Donald J. Trump, et al.*[WABE News Staff2025], where Trump was accused of inciting an insurrection in an attempt to overturn the 2020 presidential election results. The state alleged he “knowingly and wilfully joined a conspiracy to unlawfully change the outcome”[Sullivan et al.2023]. This case was brought by Fulton County District Attorney Fani Willis and hence, this topic is contextualised as ‘Legal Proceedings’.

Topic 1: This topic is more specific, as the full name of Stormy Daniels appears, indicating that it primarily concerns Trump and his 2024 hush money scandal. The inclusion of key figures, such as David Pecker, a key witness and Judge Juan Merchan, further supports this context. Therefore, this topic is labelled ‘Stormy

Daniels'.

Topic 2: This topic is focused on current events, such as Trump's handling of the Russia-Ukraine war and his controversial trade policy of imposing large tariffs on global trade. Given these themes, this topic is contextualised as 'Global Trade and Conflict'.

Topic 3: This topic primarily revolves around Kamala Harris, the Democratic contender for the 2024 election, followed by keywords such as "rally," "debate," and "today." These terms suggest the topic relates to Trump's efforts to gain political support, particularly through his campaigning activities. Thus, this topic is contextualised as 'Campaigning'.

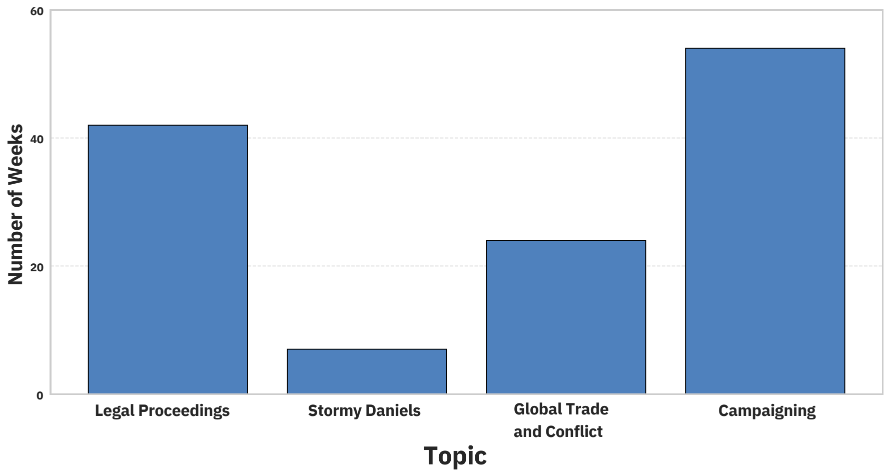

Figure 6: Weekly Topic Distribution

When we look at the distribution of these topics in Figure 6 we see that Donald Trump's Guardian headlines are mostly dominated by 'Campaigning' and 'Legal Proceedings'. While these topics constitute a significant portion of the coverage, 'Global Trade and Conflict' also emerges as a relevant theme. We can view the distribution of these topics by grouping the Guardian articles by week and assigning the most common topic discussed that week as shown in Figure 7.

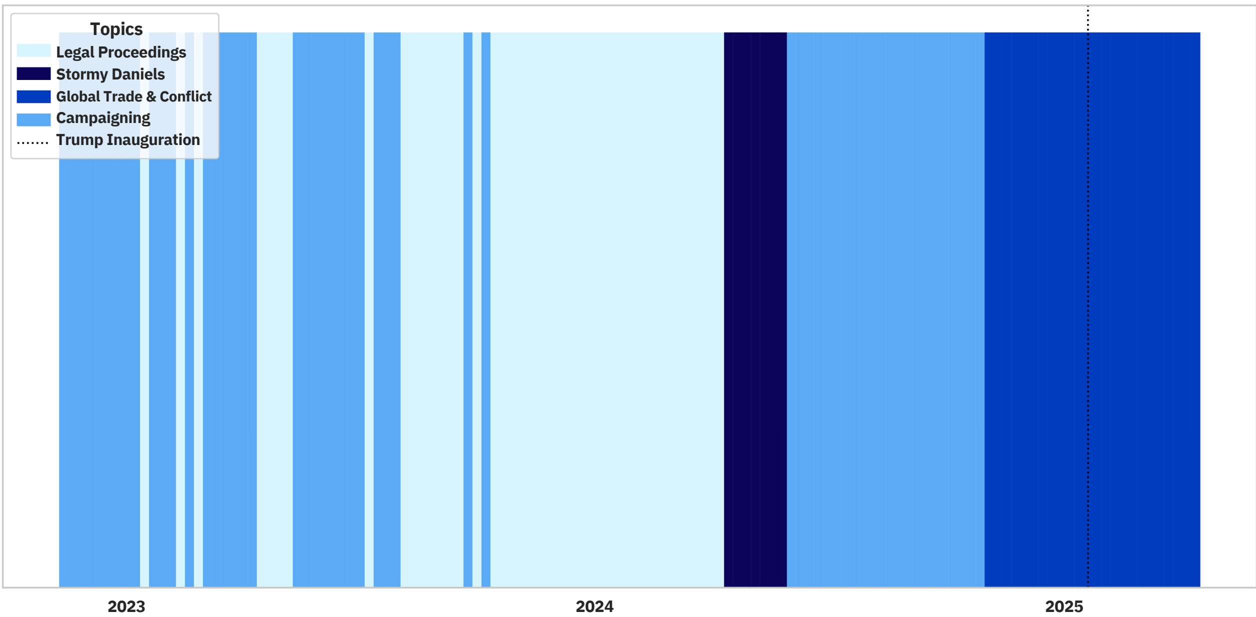

Figure 7: Weekly Topics on Trump From the Guardian

Initially the timeline of topic discussion varies between 'Campaigning' and 'Legal Proceedings' during the announcement of Trump's second run. This reflects a balance between his political developments and controversies. However, as the campaign coverage seems to become less relevant, the controversies begin to dominate the news cycle. This trend continues leading up to the 'Stormy Daniels' court case, which, although significant, was a short-lived scandal in the context of Trump-related headlines. After this, the focus shifts back to 'Campaigning'. Finally, as the 2024 election approaches, a noticeable shift occurs, with headlines becoming exclusively dominated by 'Global Trade and Conflict'. This could be a reflection of current events, notably the resurgence of attention on the Russian invasion of Ukraine. However, it also marks the introduction of Trump's tariffs and his commentary on world trade as these issues gained public attention.

4 Conclusion

4.1 Findings

Perhaps unsurprisingly, Trump is a man defined by controversy. From sexual assault allegations to attempts at overturning elections, these topics follow him like a shadow. However, his ability to generate both public outrage and support often overpowers negative press. Whether discussing tariffs and trade or his stance on Russia's invasion of Ukraine, Trump frequently redirects media attention to maintain relevance.

Despite this ability to dominate the news cycle, this strategy hasn't always worked in his favour. Initially, evoking the topic of global trade served as a useful distraction, resonating with public concerns. However, as shown in Figure 8, following a prolonged period dominated by topics on tariffs and international conflict, Trump's net favourability [YouGov2025, April 9] begins to decline, especially after his election as president. This likely reflects a growing disillusion among supporters whose expectations were not met in office.

Notably, a sharp dip in popularity during mid-2024 coincides with the Stormy Daniels court case, further illustrating the impact of legal controversies on public opinion. Interestingly, this contrasts with the preceding rise in popularity observed during Trump's indictment under Georgia's RICO statute [Times2023]. This apparent contradiction suggests that his base may rally behind him when legal action is framed as politically motivated or as an attack by the state. While allegations of a more personal nature, such as sexual assault may be less defensible, even among his supporters.

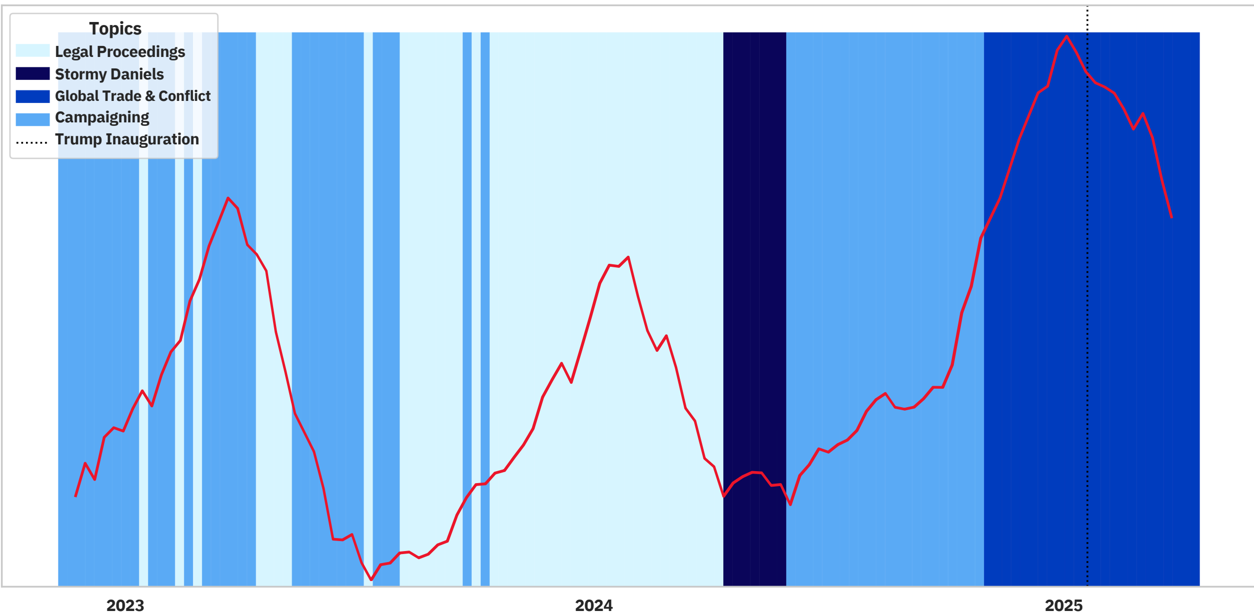

Figure 8: Weekly Topics Discussed with YouGov Opinion Poll

4.2 Hypotheses

These findings both support and challenge the initial hypotheses.
Looking at the combined graph of YouGov popularity against Guardian sentiment in Figure 5, it appears that headlines are not shaping public opinion as originally expected. Instead, there is evidence of the media responding to shifts in public opinion, albeit with a lag. This opposes the expectations set out in **H1**, but nevertheless provides a valuable insight into the media-public feedback loop.

Next, examining the Guardian's coverage of Trump, we observe a varied pattern. When Trump is not actively participating in political discourse or generating policy related news, articles frequently revert to coverage of his numerous controversies. From hush money payments and sexual assault cases to RICO charges, there is a clear abundance of content unrelated to policy. Hypothesis **H2** was therefore only partially correct: while Guardian articles largely ignore Trump's policy proposals, or lack there of. As shown in Figure 6, this emphasis shifts as the election approaches, when policy coverage becomes more prominent.

Finally, Hypothesis **H3** did not hold. Contrary to expectations, sentiment was most positive during key campaign milestones. However, this may not reflect genuine positivity, but rather a reduction in overt negativity. As coverage shifts towards campaign policies rather than personal scandals, the tone becomes more neutral; indicating not so much favourable sentiment but the absence of negative framing.

4.3 Summary

Overall, this analysis reveals a complex dynamic between media coverage, public opinion and political strategy. Rather than driving sentiment, the media appears to reflect public attitudes with a slight delay, particularly in response to controversy. Trump's ability to dominate the news cycle often through scandal, remains one of his defining political tools. However, this strategy has limits. While legal and institutional clashes can rally his base, issues like sexual assault yield more tangible drops in support. As the election approaches and the coverage turns more policy focused, sentiment becomes less extreme. This shift hints that, beneath the noise, voters may still be paying attention to what candidates are actually saying, not just what they've done.

References

Berning, C. (2023). *The polarisation of us society and its representation in the media: A linguistic analysis of selected editorials on the 2020 presidential election campaign* [Published: 27 September 2023. eBook ISBN: 978-3-658-42962-1. Series ISSN: 2625-3577]. Springer VS Wiesbaden. https://doi.org/10.1007/978-3-658-42962-1

Breitenstein, S., Kenny, J., Larner, J., Stiers, D., & Lewis-Beck, M. S. (2025). The authentic a**hole: Candidate consistency, causality, and the vote. *Polity*, *57*(2), 240–252. https://doi.org/10.1086/734478

Dada, L. S. (2023). Media's influence on voter behavior: The impact of news coverage on election outcome. *Journal of Human, Social and Political Science Research*, *2*(2). https://ssaapublications.com/index.php/sjhspsr/article/view/106

Frimpong, K., Li, A. N., Nyame, P., & Nyame, G. (2022). The impact of social media political activists on voting patterns [Published online 04 July 2020; Issue Date: June 2022]. *Political Behavior*, *44*, 599–652. https://doi.org/10.1007/s11109-020-09632-3

Hopmann, D. N., Elmelund-Præstekær, C., Albæk, E., Vliegenthart, R., & de Vreese, C. H. (2012). Party media agenda-setting: How parties influence election news coverage. *Party Politics*, *18*(2), 173–191. https://doi.org/10.1177/1354068810380097

Liu, Y. (2024). Public trust and collaborative governance: An instrumental variable approach. *Public Management Review*, *26*(2), 421–442. https://doi.org/10.1080/14719037.2022.2095003

Sullivan, A., Ax, J., & Lynch, S. N. (2023). Georgia charges trump, former advisers in 2020 election case [Accessed: 2025-04-18]. *Reuters*. https://www.reuters.com/legal/us-state-georgia-appears-set-file-charges-against-donald-trump-court-document-2023-08-14/

Takens, J., van Atteveldt, W., van Hoof, A., & Kleinnijenhuis, J. (2013). Media logic in election campaign coverage. *European Journal of Communication*, *28*(3), 277–293. https://doi.org/10.1177/0267323113478522

Times, T. N. Y. (2023). Trump indicted in georgia for election interference scheme [Accessed: 2025-04-18]. *The New York Times*. https://www.nytimes.com/2023/08/14/us/trump-georgia-rico-charges.html

WABE News Staff. (2025). *Georgia v. trump* [Accessed: 2025-04-18]. https://www.wabe.org/fulton-election-case/

YouGov. (2025, April 9). *Donald trump's declining popularity, tariffs, a third term, and the economy* [Accessed: 2025-04-18]. https://today.yougov.com/politics/articles/51986-donald-trump-declining-popularity-tariffs-third-term-the-economy-april-5-8-2025-economist-yougov-poll

A Appendix

Python Code

```
# %% [markdown]
# # Scraping headlines from the Guardian
# https://open-platform.theguardian.com/

# %%
# API Extraction
import os

# data Manipulation
import pandas as pd
import numpy as np
import random

# Visualisation
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.patches as mpatches

# Web Scraping and API Pull
import requests
from bs4 import BeautifulSoup

# Sentiment Extraction
from nltk.sentiment import SentimentIntensityAnalyzer

# Tokenisation
from nltk import word_tokenize
from nltk.corpus import stopwords

# LDA
from gensim import corpora, models
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation

# ==========================================================


# check api key is there
API_KEY = os.getenv("GUARDIAN_API_KEY")

if API_KEY:
    print("API Key Loaded Successfully")
else:
    raise SystemExit('API Key Not Found')


# %%

# define url
url = "https://content.guardianapis.com/search"

# Start date for Trump's anouncement to run again: 2022-11-15
start_date = '2022-11-15'
last_date = '2025-04-13'
```

```
params = {
    "from-date": start_date,
    "to-date": last_date,
    "api-key": API_KEY,
    "page-size": 200,
    "page": 1,
    "q": "Trump",
    "query-fields": "headline",  # Restrict search to headlines
    "show-fields": "body,headline"  # Include headline in response
}

response = requests.get(url, params=params)

if response.status_code == 200:
    Guardian_data = response.json()
    print('Scrape Successful')
else:
    print(f"Error: {response.status_code}")


# %% [markdown]
# Guardian API only scrapes 200 articles ata a time, so I have to iterate the fetch
# enough times to cover the correct number of pages to gather all the acticles
total_articles = Guardian_data['response']['total'] # ~5258
page_size = Guardian_data['response']['pageSize'] # ~200
total_pages = Guardian_data['response']['pages']  # ~27


all_headlines = []
dates = []
html_text = []

# Loop through all the pages
for page in range(1, total_pages+1):

    params["page"] = page  # Update the page number in the parameters


    response = requests.get(url, params=params)

    if response.status_code == 200:
        Guardian_data = response.json()

        # Collect all headlines from the current page
        for headline in Guardian_data['response']['results']:

            all_headlines.append(headline.get('webTitle'))
            dates.append(headline.get('webPublicationDate'))
            html_text.append(headline.get('fields').get('body'))

    else:
        print(f"Error fetching data on page {page}: {response.status_code}")

# %%

# clean text and insert into df
def clean_paragraphs(text):
```

```
    soup = BeautifulSoup(text, 'html.parser')
    paragraphs = soup.find_all('p')
    return " ".join(p.text for p in paragraphs)




data = pd.DataFrame({
    'headline': all_headlines,
    'date': pd.to_datetime(dates).dt.tz_localize(None),#.floor('d'),
    'text': html_text
})


data['week_start'] = data['date'].dt.to_period('W').dt.start_time


data['text'] = data['text'].apply(clean_paragraphs)

# %%
headline_counts = data.groupby('week_start')['headline'].count().reset_index()


plt.figure(figsize=(20, 10))
plt.bar(headline_counts['week_start'], headline_counts['headline'], color='#4F81BD',
        width=7,        edgecolor='black')

plt.xlabel('Year', labelpad=25)
plt.ylabel('Number of Headlines')
plt.tight_layout()
plt.grid(False)
plt.yticks([0, 50, 100, 150])
tick_dates = pd.to_datetime(['2023-01', '2024-01', '2025-01'])

# Apply to x-axis
plt.xticks(ticks=tick_dates, labels=['2023', '2024', '2025'])

plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.axvline(pd.to_datetime('2025-01-20'), linestyle='--', color='red',
            label='Trump Inauguration')
plt.legend()
plt.savefig('../Plots/Weekly_Occurance.pdf', format='pdf',
            bbox_inches='tight')
plt.show()

# make the y axis only show 0,50,100 and 150


# %% [markdown]
# # Quick Sentiment analysis

# %%
# ~ 1:30 mins

# Initialise the sentiment analyser
sia = SentimentIntensityAnalyzer()

# Apply to the text column
```

```
data['sentiment_scores'] = data['text'].apply(sia.polarity_scores)

# Optionally split into separate columns
sentiment_df = data['sentiment_scores'].apply(pd.Series)
data = pd.concat([data, sentiment_df], axis=1)

display(data)


# %% [markdown]
# ### Visualising the Sentiment

# %%
data.sort_values('date', ascending=False, inplace=True)

plt.figure(figsize=(14, 6))
plt.bar(data['date'], data['neg'], color='crimson', width=1.5)
plt.title('Negative Sentiment Over Time', fontsize=16)
plt.xlabel('Date', fontsize=12)
plt.ylabel('Negative Sentiment Score', fontsize=12)
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.6)

plt.tight_layout()
plt.show()

# group sentiment scores by week
weekly_sentiment = data.groupby('week_start')[['neg', 'pos']].mean().reset_index()

# plot
weekly_sentiment.plot(x='week_start', kind='bar', figsize=(12, 6))
plt.title('Weekly Average Positive vs Negative Sentiment')
plt.xlabel('Week Starting')
plt.ylabel('Average Sentiment Score')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# line plot of trends
plt.figure(figsize=(12, 6))
plt.plot(weekly_sentiment['week_start'], weekly_sentiment['neg'],
         label='Negative',
         color='red')
plt.plot(weekly_sentiment['week_start'], weekly_sentiment['pos'],
         label='Positive',
         color='green')
plt.title('Weekly Sentiment Trend Over Time')
plt.xlabel('Week Starting')
plt.ylabel('Average Sentiment Score')
plt.legend()
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()


weekly_sentiment['net_sentiment'] = weekly_sentiment['pos'] - weekly_sentiment['neg']

plt.figure(figsize=(12, 6))
```

```
plt.plot(weekly_sentiment['week_start'], weekly_sentiment['net_sentiment'], marker='o')
plt.axhline(0, linestyle='--', color='grey')
plt.title('Weekly Net Sentiment (Positive - Negative)')
plt.xlabel('Week Starting')
plt.ylabel('Net Sentiment')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()



# %%
weekly_sentiment['net_sentiment'] = weekly_sentiment['pos'] - weekly_sentiment['neg']

# Apply a moving average for smoothing (you can adjust the window size)
weekly_sentiment['smoothed_sentiment'] = weekly_sentiment['net_sentiment'].rolling(window=

plt.figure(figsize=(12, 6))
plt.plot(weekly_sentiment['week_start'], weekly_sentiment['smoothed_sentiment'],
         color='#4169E1',
         label='Sentiment')  # Smoothed line
plt.plot(weekly_sentiment['week_start'], weekly_sentiment['net_sentiment'],
         alpha=0.3,
         label='Net Favourability',
         color='grey')

plt.axhline(0, linestyle='--', color='#36454F')
plt.axvline(pd.to_datetime('2025-01-20'), linestyle='-', color='#36454F',
            label='Trump Inauguration')


plt.xlabel('Year')
plt.ylabel('Net Sentiment')
# plt.xticks(rotation=45)
plt.legend()
plt.tight_layout()
plt.grid(False)
plt.xticks(ticks=tick_dates, labels=['2023', '2024', '2025'])

plt.savefig('../Plots/Weekly_Sentiment.pdf', format='pdf', bbox_inches='tight')


plt.show()


# %% [markdown]
# # YouGov Data
# https://today.yougov.com/politics/articles/51986-donald-trump-declining-popularity-tarif

# %%
YouGov = pd.read_csv('../data/YouGov.csv')
YouGov.columns = ['date', 'favorable', 'unfavorable']

YouGov['date'] = pd.to_datetime(YouGov['date'])

# Remove timezone info from both if present
YouGov['date'] = pd.to_datetime(YouGov['date']).dt.tz_localize(None)
```

```
# Then filter
min_date = data['date'].min()
YouGov = YouGov[YouGov['date'] >= min_date]


# %%
# Compute net favourability
YouGov['net_favourable'] = YouGov['favorable'] - YouGov['unfavorable']

# Apply a rolling average to smooth the net favourability
YouGov['net_favourable_smoothed'] = YouGov['net_favourable'].rolling(window=10, center=Tru

# Plotting
plt.figure(figsize=(12, 6))
plt.plot(YouGov['date'], YouGov['net_favourable_smoothed'], label='Smoothed Net Favourabil
         color='#eb152a')
plt.plot(YouGov['date'], YouGov['net_favourable'], alpha=0.3, label='Net Favourability', c
plt.axhline(0, linestyle='--', color='#36454F')
plt.axvline(pd.to_datetime('2025-01-20'), linestyle='-', color='#36454F', label='Trump Ina


plt.xlabel('Year')
plt.ylabel('Net Favourability (Favourable - Unfavourable)')
plt.xticks(ticks=tick_dates, labels=['2023', '2024', '2025'])
plt.legend()
plt.savefig('../Plots/yougov_opinion.pdf', format='pdf', bbox_inches='tight')

plt.show()




# %%
fig, ax1 = plt.subplots()

# Plot the first line with the primary y-axis
ax1.plot(weekly_sentiment['week_start'], weekly_sentiment['smoothed_sentiment'], color='#4
         label='Guardian')
ax1.axvline(pd.to_datetime('2025-01-20'), linestyle='--', color='#36454F', label='Trump In

ax1.set_xlabel('Year')
ax1.set_ylabel('Net Favourability')
ax1.set_yticks([])

# Create a second y-axis
ax2 = ax1.twinx()

# Plot the second line with the secondary y-axis
ax2.plot(YouGov['date'], YouGov['net_favourable_smoothed'], color='#eb152a', label='YouGov
ax2.set_yticks([])

# Show the plot
# plt.title('Sentiment and Net Favourability')
fig.tight_layout()  # To prevent overlap of labels
plt.xticks(ticks=tick_dates, labels=['2023', '2024', '2025'])

# Combine both legends
handles, labels = ax1.get_legend_handles_labels()  # Get the handles and labels from ax1
```

```
handles2, labels2 = ax2.get_legend_handles_labels()  # Get the handles and labels from ax2
ax1.legend(handles=handles + handles2, labels=labels + labels2)  # Combine the handles and



plt.savefig('../Plots/yougov_guardian.pdf', format='pdf', bbox_inches='tight')

plt.show()

# %% [markdown]
# # Separating the Key Words

# %%
# ~
# Load stopwords
stop_words = set(stopwords.words('english'))
common_words = ['said', 'us', 'new', 'last', 'first', 'cohen', 'say', 'gmt', 'bst', 'one',
                'says', 'read', 'full', 'story','updated', 'going', 'people', 'donald', 'h

def clean_and_tag(text):
    # Tokenize and lowercase
    tokens = word_tokenize(text.lower())

    # Filter out stopwords, non-alpha, and common words
    words = [word for word in tokens if word.isalpha() and word not in stop_words and word

    # return words, descriptive_words, nouns
    return words
## Apply the function
# data[['clean_tokens', 'descriptive_words', 'nouns']] = data['text'].apply(
    # lambda x: pd.Series(clean_and_tag(x))
# )
data['tokens'] = data['text'].apply(lambda x: clean_and_tag(x))


# %% [markdown]
# # Grouping the Articles by Week

# %%
data['grouped_tokens'] = data['tokens'].apply(lambda x: ',-'.join(x))

weekly_tokens = data.groupby('week_start')['grouped_tokens'].agg(',-'.join).reset_index()


weekly_tokens


# %% [markdown]
# # LDA
#

# %% [markdown]
# ### Initialising the LDA

# %%
SEED = 15

random.seed(SEED)
```

```
np.random.seed(SEED)


# ~ 40 seconds

# Use CountVectorizer, not TF-IDF
vectorizer = CountVectorizer(
    max_df=0.98,     # Ignore terms that appear in >95% of weeks
    min_df=2,        # Ignore rare terms
    stop_words='english'
)

X = vectorizer.fit_transform(weekly_tokens['grouped_tokens'])


# %%



# Choose a smaller number of topics, depending on how many weeks you have
lda = LatentDirichletAllocation(
    n_components=4,    # Try 3-6 and compare coherence
    random_state=SEED,
    max_iter=20,
    learning_method='batch'
)

lda.fit(X)

# Assign topic distribution
topic_distributions = lda.transform(X)

# Assign dominant topic per week
weekly_tokens['topic'] = topic_distributions.argmax(axis=1)



# %%
weekly_tokens['topic'].value_counts()

# %%
# Extract the vocabulary
words = vectorizer.get_feature_names_out()

# Loop through each topic
for i, topic in enumerate(lda.components_):
    # Get the top 15 words for each topic
    top_words = [words[i] for i in topic.argsort()[-5:][::-1]]

    # Print the topic number and its top words
    print(f'Topic {i}:', ", ".join(top_words))




# %%
legend_labels = {
    0: "Legal Proceedings",
```

```
    1: "Stormy Daniels",
    2: "Global Trade and Conflict",
    3: "Campaigning"
}


# %%
# count the occurrences of each topic
topic_counts = weekly_tokens['topic'].value_counts().sort_index()

# plot
plt.figure(figsize=(20, 10))
bars = plt.bar(topic_counts.index, topic_counts.values, color='#4F81BD', edgecolor='black'
plt.xlabel('Topic')
plt.ylabel('Number of Weeks')
plt.xticks(topic_counts.index)  # Ensure all topic IDs are shown
plt.grid(False)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.yticks([0, 20, 40, 60])

topic_ids = list(legend_labels.keys())
topic_names = [legend_labels[i] for i in topic_ids]
plt.xticks(ticks=topic_ids, labels=topic_names)

plt.tight_layout()
plt.savefig('../Plots/Topic_Distribution_Per_Week.pdf', format='pdf', bbox_inches='tight')
plt.show()


# %%
weekly_tokens['val'] = 1

# Display the updated dataframe



# %%

# Define palette and legend labels
palette = ['#dbf6ff',
           '#0c085c',
           '#043cbe',
           '#5eaef9']


# Create the plot
plt.figure(figsize=(20, 10))

# Bar plot
plt.bar(weekly_tokens['week_start'],
        weekly_tokens['val'],
        color=[palette[i] for i in weekly_tokens['topic']],
        width=8,
        align='center',

        edgecolor=[palette[i] for i in weekly_tokens['topic']])

# Add Inauguration line
```

```
inauguration_line = plt.axvline(pd.to_datetime('2025-01-20'),
                                color='black',
                                linestyle='dotted',
                                linewidth=2,
                                label='Trump Inauguration')

# Create custom legend for topic categories
patches = [mpatches.Patch(color=palette[i], label=legend_labels[i]) for i in range(len(leg

plt.legend(handles=patches + [inauguration_line], title='Topics', loc='upper left')

# Additional plot settings
# plt.xticks(rotation=45)
plt.yticks([])
plt.grid(False)
plt.xticks(ticks=tick_dates, labels=['2023', '2024', '2025'])

plt.tight_layout()
plt.savefig('../Plots/Weekly_Topics.pdf', format='pdf', bbox_inches='tight')
plt.show()


# %%
#  net favourability
YouGov['net_favourable'] = YouGov['favorable'] - YouGov['unfavorable']
YouGov['net_favourable_smoothed'] = YouGov['net_favourable'].rolling(window=10, center=Tru

# plot ONLY the smoothed red line
plt.figure(figsize=(12, 6))
plt.plot(YouGov['date'], YouGov['net_favourable_smoothed'], color='#eb152a')
plt.axis('off')
plt.savefig('../Plots/yougov_redline_only.pdf', format='pdf', bbox_inches='tight', pad_inc
plt.close()
```