Opleiding Informatica

Analyzing Communication on Bluesky:

Sentiment, Emotion, and Toxicity in Dutch Political Discourse

Olaf Tijs

Supervisors:
Dr. A. Saxena & F. Corriera

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

19/08/2025

Abstract

This thesis investigates gender-based differences in sentiment, emotion, toxicity, and topical engagement within Dutch political discourse on the decentralized social media platform Bluesky. Motivated by the platform's growing role in shaping online political discussion, the study applies state-of-the-art NLP methods to a dataset collected using the ATProto API. Posts were embedded with TwHIN-BERT, after which K-means clustering was applied to identify thematic topics. Six distinct topics were found, with varying user diversity, expressed sentiment, and user demographics. Sentiment and emotion analysis revealed that the majority of posts expressed positive sentiment, while *joy* and *anticipation* were the most prevalent emotions. Female users expressed more negative sentiment and more joy and sadness, while male users more often expressed anger, disgust, and anticipation. Toxicity levels were quite low, though male-authored posts scored higher on average. These findings highlight the influence that different user demographics and languages can have on online discourse.

Contents

1 Introduction

The rise of social media platforms in the digital age has transformed how we communicate, share opinions, and interact with each other. On a plethora of platforms, enormous volumes of content are created every day. Thorough research has been done on the content posted on several of these social media platforms, including Twitter, Facebook, and Reddit [23]. As new platforms emerge, understanding the sentiment expressed by users on such platforms can be valuable to businesses, researchers, and policymakers alike [24]. One of those emerging platforms is **Bluesky**, a decentralized social network designed to enable open and transparent communication. In contrast to its peers like X (formerly known as Twitter), it is open source and decentralized, setting it apart from the traditional centralized platforms by moving the control from a corporation to its community, which promotes transparency [27].

Understanding how these structural differences influence the nature of online conversation can be vital, especially as decentralized platforms continue to grow. While sentiment analysis has been extensively applied to platforms such as Twitter, Facebook, and Reddit, very little research has been done on Bluesky, and even less on the subject of politics specifically. The research so far has shown Bluesky users mostly post left-leaning news sources [39], that most users tend to share the same stance on political issues [42] and that left-leaning users tend to exhibit more toxicity than centre- or right-leaning users [34]. Analysing how people engage politically on the platform can offer valuable insights into democracy and public opinion.

This thesis contributes to this emerging field by analysing Bluesky posts related to Dutch politics using sentiment and emotion analysis and topic modelling. The Netherlands has a distinctive political landscape, characterized by a political system that is one of the most proportional systems in the world, a society that only recently evolved from pillarization, and the "poldermodel" based on compromises [8]. This diversity is reflected in online conversations, where citizens engage in political discussion, critique policies, and share opinions on current events. Analysing sentiment and emotion in this setting can yield insights into how Dutch political discourse develops in a decentralized online environment.

This thesis aims to answer the following question:

How is sentiment expressed in Bluesky posts about Dutch politics?

In the process, we will answer the following sub-questions:

1. How do sentiment and emotion differ between languages and genders on Bluesky posts about Dutch politics?

2. What is the difference in the presence of toxicity, threats, insults, and identity attacks on Bluesky posts about Dutch politics

3. What are the main topics being discussed related to Dutch politics on Bluesky?

4. How do topics differ in gender and language composition, sentiment, emotion, and toxicity?

To address this question, this research will collect a corpus of Bluesky posts related to Dutch politics, apply sentiment and emotion analysis techniques along with topic modelling, and analyse the sentiment expressed over time. To achieve this, a dataset of Bluesky posts will be collected using the ATProto API [4]. This dataset will contain all posts related to Dutch politics, collected using corresponding hashtags. The gender of users will be inferred using [10]. Posts will be filtered by political keywords and analysed using RoBERTje [18] and RoBERTa [38] for predicting Dutch and English sentiment, respectively, TweetNLP [14] for emotion analysis [14], and Google perspective [30] for perspective analysis. Finally, Posts will be assigned to clusters using K-means clustering.

1.1 Thesis overview

This thesis is structured as follows: Section 2 discusses related work and provides an overview of theory and techniques in the field. Section 3 shows the methods we used in our research. Section 4 interprets our discussion of our findings, limitations of the study, and potential further research opportunities. 5 concludes the thesis with our findings.

2 Related Work

This section will review previous research done on sentiment and emotion analysis, as well as topic modelling. It will look at studies on other platforms like X, as well as the limited work on Bluesky.

2.1 Sentiment Analysis on Social Media Platforms

Sentiment analysis [43] has been extensively applied to content from popular centralized social media platforms such as Twitter, Facebook, and Reddit. These platforms provide large volumes of user-generated public data and have become standard benchmarks for studying online discourse. Sidarenka [46] conducted sentiment analysis on German-language tweets, highlighting challenges that are posed by informal language, abbreviations, and slang. The study compares lexicon-based methods to machine learning techniques, concluding that the latter generally outperforms rule-based methods in handling noisy social media text. This aligns with a broader trend in the field, where machine- and deep-learning models have increasingly outperformed rule-based approaches.

Recent advances in deep learning, especially transformer-based models such as BERT [19], have further improved sentiment analysis. Fine-tuned BERT models have been shown to outperform traditional models on various occasions, for example on Twitter [47]. For more low-resource languages than English, the multilingual mBERT has shown a lot of potential to be fine-tuned on most languages, even those with as few resources as Faeroese [33]. Both de Vries et al. [17] and Delobelle et al. [18] used mBERT to train monolingual Dutch models, BERTje and robBERT, respectively. Both of these outperform mBERT, with the downside that neither of them is multilingual. De Bruyne et al. compared these two, finding that robBERT outperformed BERTje in almost all cases [16]. These studies mark an important starting point for sentiment analysis on Dutch text, but there are still plenty of domains to be explored.

2.2 Sentiment Analysis in Political Context

Sentiment analysis has also been widely used to study political discourse. In the Dutch context, Gielens et al. [21] examined Twitter discourse on Universal Basic Income (UBI). They found that while left-wing parties tend to support UBI more, the overall polarization between opponents and proponents is limited. The study highlights the importance of sentiment analysis in understanding ideological divides on policies.

Schumacher et al looked at a more traditional method of political communication [45], analysing speeches from party congresses in the Netherlands, finding party leaders use more negative language compared to ministers and other party members, supporting earlier international findings [40, 15, 28]. However, they did not replicate results suggesting moderate parties use more positive language [15] or that there was a strong increase in the use of positive sentiment over time [40]. This suggests that sentiment patterns in Dutch political discourse may not align with broader international trends and proves the need for further research.

From an international perspective, Zhang et al. [53] analysed sentiment during elections on Twitter and found that negative sentiment dominates election discourse on Twitter. They also noted that negative content spread more widely and faster than neutral or positive messages and that female candidates were disproportionately targeted with negative language. These studies prove that sentiment is a valuable tool for examining political sentiment and can lead to important discoveries about sentiment expression trends.

2.3 Gender-Based Variation in Expressed Sentiment

Gender plays a notable role in how sentiment is expressed, with plenty of recent studies examining differences in linguistic patterns and interaction dynamics across male and female users.

Thelwall et al. [49] found female users expressed more positive sentiment, although there was no significant difference in negative sentiment. Panchendrarajan et al. studied sentiments on Twitter communication and observed similar results [36]. Macedo and Saxena studied sentiment for Soccer communication in English and Portuguese and observed that female express stronger sentiments whenever any event happens [31].

Rollero et al. [41] investigated gender differences in online social networks for Italian users. They found that men tend to post more often than women on social media. In the domain of politics, these findings were backed up when analysing posts by German journalists on Twitter [51] and posts by Spanish MPs [22].

Similarly, research shows posts by women are also shared a lot less. Manzano et al. Found that female influencers had significantly less influence on political Twitter networks than their male counterparts, being retweeted less [32]. Witzenberger et al. [51] showed that tweets by male journalists were retweeted more.

These findings suggest a pattern where men are more active in posting, reinforcing the need to consider gender dynamics in online political communication.

2.4 Sentiment on Bluesky

Although Bluesky is a relatively new platform, emerging literature has begun to explore its user behaviour, content, and dynamics.

Reddy et al. [39] and Nogara [34] both found that Bluesky users tend to post mostly left-leaning news sources, with the latter also finding more left-leaning users tended to show more toxicity compared to centre- and right-leaning users. Salloum et al. [42] found that most users tend to hold similar views on politics. These findings suggest that Bluesky's user base is currently a specific subgroup, suggesting a possible difference in expression compared to traditional social media and offering opportunities for finding meaningful patterns in future research.

2.5 Emotion Expression and Toxicity

Research about the expression of emotion across different platforms, genders, and other groups can provide valuable insights into the communication in online spaces. Kivran et al. [26] found female

Twitter users expressed more positive emotions, especially in female-to-female interactions. Several studies found women express more emotion online, particularly positive emotions [37, 35, 20]. When looking at toxicity, several studies have found men to exhibit more of it than women [31, 48]. When comparing platforms, Waterloo et al. [50] found significant differences in emotion expression between platforms, with Twitter having a relatively high amount of negative and a low amount of positive emotions compared to other platforms. Avelle et al. [11] found relative consistency in the amount of toxicity across platforms, and that toxicity attracts more interaction rather than deterring it.

Previous research indicates that different genders express different emotions, and that the emotions expressed differ across platforms. This indicates valuable results can be achieved when looking at newer platforms.

2.6 Shortcomings and Biases of Sentiment Analysis Models

Despite their widespread use on social media platforms, sentiment analysis models are known to exhibit systemic biases. One such bias was shown by Thelwall [49], who conducted a study that demonstrated gender bias in sentiment analysis, using TripAdvisor reviews. The study found that sentiment expressed by female authors was detected more accurately than their male counterparts. Kiritchenko et al. [25] found that there was a significant difference in the sentiment assigned to text about male and female subjects, and between European and African names. Recognizing and accounting for such biases is essential to ensure the quality and fairness of sentiment analysis.

2.7 Research Gap

While sentiment analysis is well explored in centralized platforms such as Twitter and Facebook, there are limited studies on decentralized platforms like Bluesky. This research aims to address this gap by analysing the sentiment on Bluesky, considering its unique decentralized nature and the implications for content moderation, user engagement, and online discourse.

In addition to the aforementioned gap in platform coverage, there is also a notable linguistic gap. Although sentiment analysis in Dutch has picked up steam in recent years, partly due to the monolingual models BERTje and robBERT, most of the work has been limited to domains such as product reviews and news articles or centralized social media platforms. To date, there is no study on Dutch-language sentiment on decentralized platforms like Bluesky. This study seeks to address both the platform and language gaps by performing Dutch-language sentiment analysis specifically on Bluesky. In doing so, it contributes to the under-explored section of decentralized communication and medium-resource languages in sentiment analysis, offering new insights into expressed sentiment online.

3 Data Analysis and Insights

This study aims to analyse how sentiment is expressed in Bluesky posts about Dutch politics. To answer this question, it employs a computational pipeline that involves collecting, processing, and classifying the sentiment of user-generated content from Bluesky.

3.1 Data collection

For this study, a dataset was collected using the ATProto API [4], which consists of Dutch and English posts containing any of 467 political hashtags. The list of hashtags was compiled through an iterative expansion process. Starting with a list of hashtags containing the names of Dutch governing parties, we retrieved all associated posts and extracted all the hashtags these posts contained. All non-political hashtags were removed, and this process was repeated three times. Hashtags were used to ensure relevance to political discourse, with government party names giving the study a starting point that reflects the most important actors in Dutch politics.

Once the final list of hashtags was established, posts containing any of these hashtags were collected over a three-month period, from March 17, 2025, to June 17, 2025. Language filtering was performed based on metadata provided by Bluesky, retaining only Dutch and English posts.

For each post, both the post object and the corresponding author's user object were retrieved. In total, the dataset contains 38,824 posts made by 7,229 unique users.

	Dutch	English	Both	Total
Number of posts	25015 (64.4%)	13588 (35.0%)	221 (0.6%)	38824
Number of unique users	2641 (36.5%)	4214 (58.3%)	374 (5.2%)	7229

Table 1: Statistics of the Bluesky Data. 'Both' refers to posts that contain both languages or users who made posts in both languages.

3.2 Ethical statement

This study relies exclusively on publicly available data collected from the Bluesky platform via the ATProto API. All post and user metadata included in the dataset were accessible without authentication, and no private content was accessed during the study.

To minimize privacy concerns, the dataset is stored without any information that can be used to identify users from the username object, to avoid privacy infringement.

Given the public nature of Bluesky and the fact that users voluntarily share content in a public forum, this study aligns with current ethical standards for observational research using social media data [9].

3.3 Data processing

To construct a dataset for further analysis, the following fields were extracted from each post object and the corresponding user object:

1. **DID**: The Decentralized Identifier (DID) of the post's author. This is a unique personal identifier that remains consistent even if the display name is changed.

2. **Displayname**: The display name of the post's author. If the user has changed their display name during the collection period, the most recent display name was used for all their posts. If the user never set a display name, their handle was used instead.

3. **Datetime**: The timestamp of the post's creation, formatted into ISO 8601

4. **PostText**: The full textual content of the post, including mentions, links, and hashtags.

5. **PostLength**: The number of words in the post.

6. **Hashtags**: A list of all the Hashtags in the posts.

7. **Emoticons**: A list of all emoticons present in the post, extracted using regular expressions.

8. **Language**: A list of languages detected in the post, based on metadata provided by Bluesky. Regional variants (e.g.,B ritish or American English) were standardized to a generic label (e.g., English)

After collecting the complete post objects, duplicate entries were identified and removed. Duplicates were defined as posts with identical PostText values.

3.3.1 Gender Inference

To infer the gender of users in our dataset, we applied the genderize function of NameAPI [10] to users' display names. For users who did not set a display name, their handle was used instead.

This process returns three values, each of which was added to the dataset:

- **Gender**: The predicted gender. Possible values include *male*, *female*, *neutral* (the name may refer to both genders—see *maleProportion*), and *unknown*. Although the API actually distinguishes between three distinct values of unknown predictions, these were all grouped under *unknown*, as the distinctions are not relevant for our scope.

- **maleProportion**: The proportion of people with that name who are male. This value is only relevant for names classified as *neutral*, otherwise it is left empty.

- **Confidence**: The model's confidence in its prediction, ranging from 0 to 1, with 1 indicating maximum confidence.

The API returned gender predictions for 3,256 users who collectively made 12,638 posts. This represents 45% of users and 32.5% of all posts in the dataset.

	Male	Female	Unknown	Total
Users	2,163 (29.9%)	1,093 (15.1%)	3,973 (55.0%)	7,229
Posts made by	9,057 (23.3%)	3,581 (9.2%)	26,186(57.5%)	38,824

Table 2: Number of unique users and posts by predicted gender

3.4 Sentiment Analysis

Next, we perform sentiment analysis on the collected dataset [44]. To analyse the sentiment of post content, two BERT-based models were used. For Dutch posts, we used RobBERT [18], which has been shown to outperform BERTje [17], the other major Dutch sentiment analysis model, in nearly all cases [16]. For English, we used RoBERTa-base [38].

To quantify sentiment to allow for statistical analysis, we mapped the predicted labels to numerical values: *negative* to -1, *neutral* to 0, and *positive* to 1. Posts including both Dutch and English are processed by RoBERTa. The results of the sentiment analysis are visualized in figures 1 through 4.

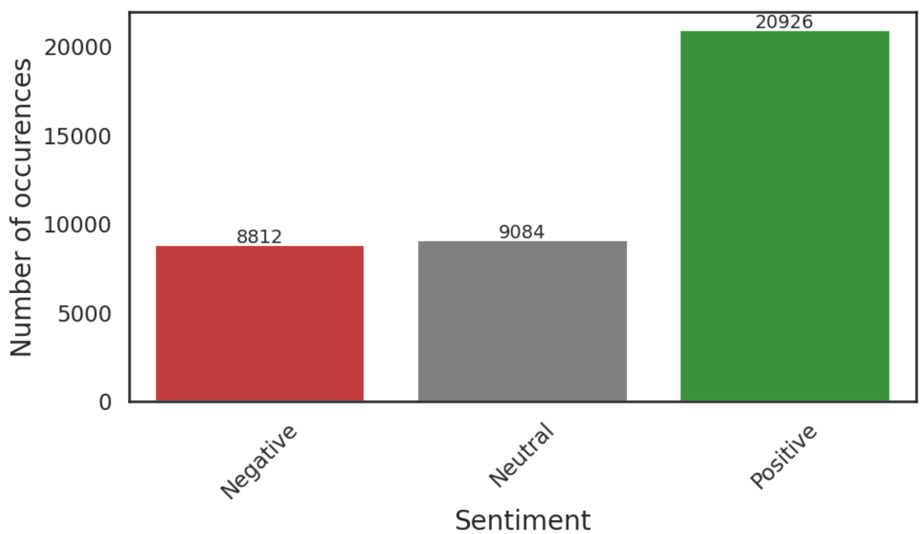

Figure 1: Sentiment distribution across all posts.

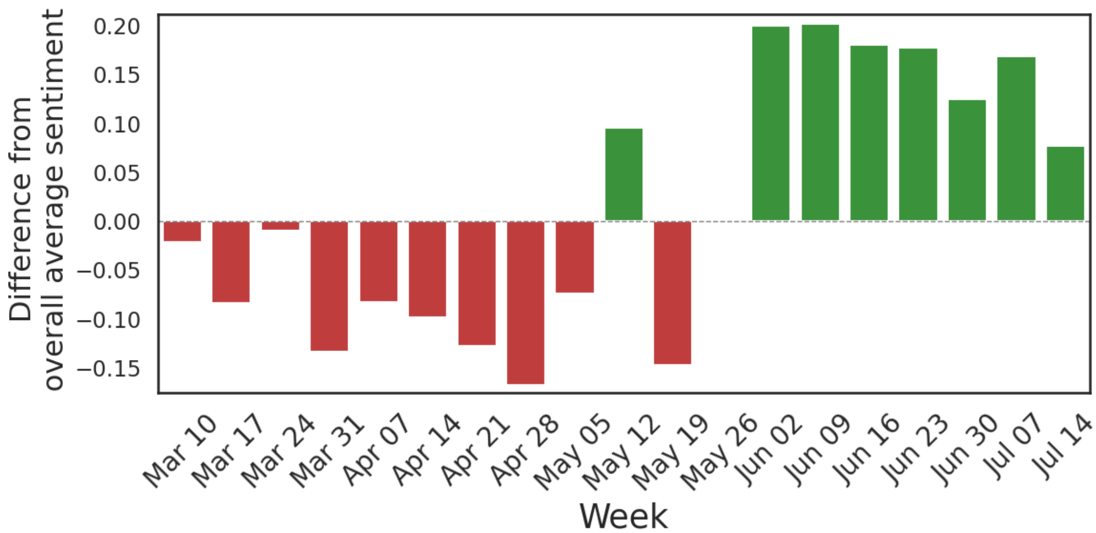

Figure 2: Weekly average sentiment over time.

Figure 1 shows that more than half of the posts in the dataset are classified as positive.

Figure 2 reveals a noticeable shift in sentiment beginning in the week of June 2nd. The average sentiment score increases from 0.231 before this date to 0.479 after. Post volume also decreases during this period, from an average of 326 posts per day to 294.

This change coincides with the collapse of the Dutch cabinet on June 3rd [5], which may be an explanation for both the sudden increase in sentiment and the drop in post volume.

To test whether the observed increase in sentiment after June 2nd was statistically significant, a Mann–Whitney U test was performed. The result ($p < .001$) confirms a highly significant difference in sentiment before and after the collapse of the Dutch cabinet, though this does not establish causation.

Figure 3 shows that posts in both languages express negative sentiment at similar levels, but Dutch posts are more likely to express positive sentiment.
Figure 4 indicates that male users tend to post more neutral content, whereas female users are more likely to express negative sentiment. This could be explained by earlier findings that suggest sentiment expressed by women is more accurately detected by sentiment analysis models [49, 29].

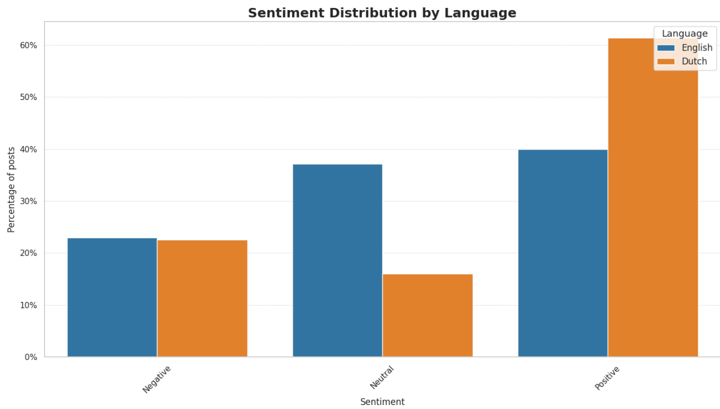

Figure 3: Sentiment distribution by language

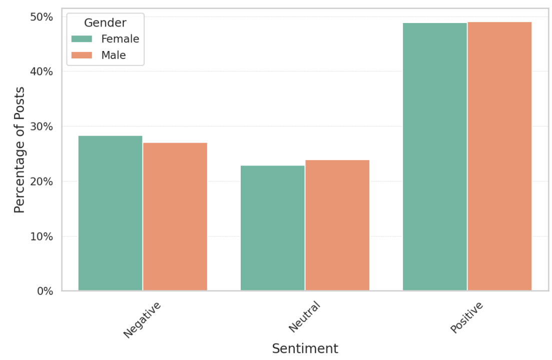

Figure 4: Sentiment distribution by predicted gender

3.5 Emotion Analysis

For emotion classification, we used the TweetNLP model [14], which assigns scores (ranging from 0 to 1) across eight emotion categories *joy*, *trust*, *fear*, *surprise*, *sadness*, *disgust*, *anger*, and *anticipation*.

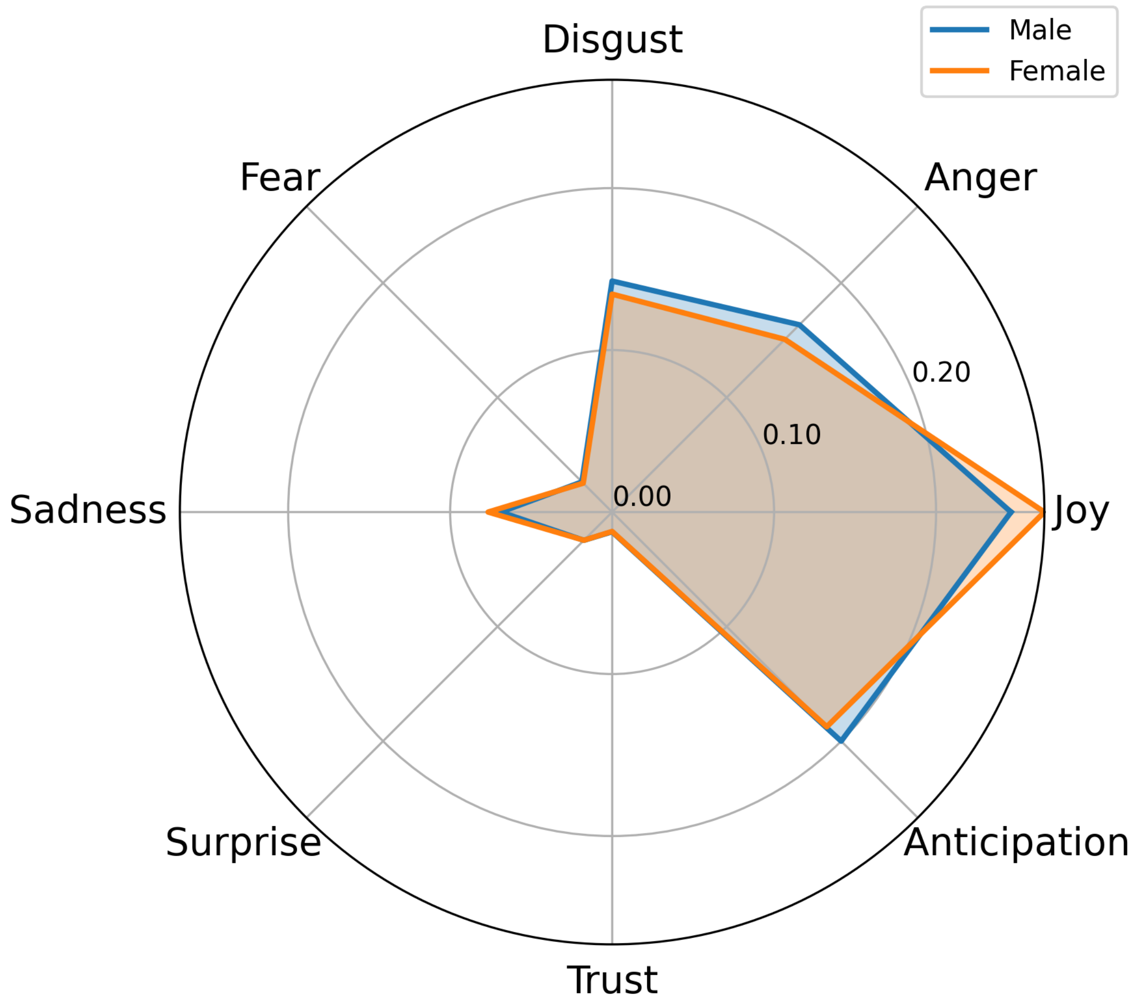

Figure 5: emotions of posts made by male and female posters

Figure 5 illustrates and compares the average emotion scores in posts by male and female users. *Anticipation* and *joy* are the most prominent emotions across the dataset, while *fear*, *surprise*, and *trust* are the least common. Comparisons between genders reveal that male users express 8% more anger, 7% more anticipation and 6% more disgust than female users. Conversely, female users express 8% more joy and 11% more sadness. Other emotions are expressed on a relatively equal scale by both genders.

To explore the relationship between emotion and sentiment, we conducted an Ordinary Least Squares (OLS) regression using the emotion scores as predictors and sentiment (mapped from -1 to 1) as the dependent variable. Prior to fitting the model, all emotion features were normalized using Z-score normalization. As a result, each regression coefficient can be interpreted as the expected change in sentiment score associated with a one standard deviation increase in the corresponding emotion score.

Variable	**Coef.**	**P-value**	**Confidence Interval (95%)**	
			Lower	**Upper**
Intercept	0.3284	0.000	0.320	0.336
Joy	-0.0217	0.157	-0.052	0.008
Anger	-0.3310	0.000	-0.360	-0.302
Fear	-0.1130	0.000	-0.126	-0.100
Sadness	-0.0832	0.000	-0.098	-0.068
Surprise	0.0714	0.000	0.059	0.083
Trust	-0.1217	0.000	-0.145	-0.099
Disgust	0.0540	0.000	0.031	0.077
Anticipation	0.0023	0.827	-0.018	0.023

Table 3: OLS Regression Results

The regression results in Table 3 indicate that several emotions significantly predict sentiment, though the explanatory power is modest (R^2=0.095). This suggests that the emotion scores only explain about 9.5% of the variance in sentiment. The *Intercept* of 0.3284 represents the baseline predicted sentiment if all independent variables are zero. In other words, posts with no strong emotion present are generally positive. Among the emotions, *anger, fear,* and *sadness* are significantly associated with lower sentiment scores, indicating a strong correlation with more negative sentiment. Conversely, *surprise* has a significant positive impact on sentiment, suggesting posts expressing surprise are generally more positive. Interestingly. *Trust*, generally considered a positive emotion, shows a significant negative association with sentiment. However, this should be interpreted with caution: The mean score is 0.012, the lowest of any emotion, with the highest score only being 0.079, indicating that this result may be influenced by noise due to the limited variation in this emotion's scores. Another surprising finding is that *disgust* exhibits a small but significant positive impact on sentiment, despite its usual negative connotation. Unlike *trust*, *disgust* is more prevalent in the dataset, with a mean score of 0.14 Finally, *joy* and *anticipation* do not show a statistically significant relation with sentiment.

3.6 Perspective Analysis

To analyse toxicity, insults, threats, and identity attacks, this study utilizes the Google Perspective API. Each of these aspects gets scored on a scale from 0 to 1, where 1 is the highest score. Results are shown in Table 4 Figure 6 to 9. They show a difference between genders: the distribution of male scores is more right-skewed for all perspectives. A Kolmogorov–Smirnov (K-S) test for significance shows that there is a significant difference ($p < 0.05$) between the score distributions for *Toxicity*, *Insult*, and *Identity attack*, but not *Threat*.

	Perspective	**Male**	**Female**	**P-value**
	Toxicity	0.154 ± 0.168	0.152 ± 0.170	0.021
	Insult	0.111 ± 0.163	0.108 ± 0.160	0.026
Post-wise Avg. perspective	Threat	0.027 ± 0.065	0.027 ± 0.066	0.173
	Identity Attack	0.073 ± 0.125	0.072 ± 0.124	0.042

Table 4: Average perspective API score with standard deviation and p-value of K-S test

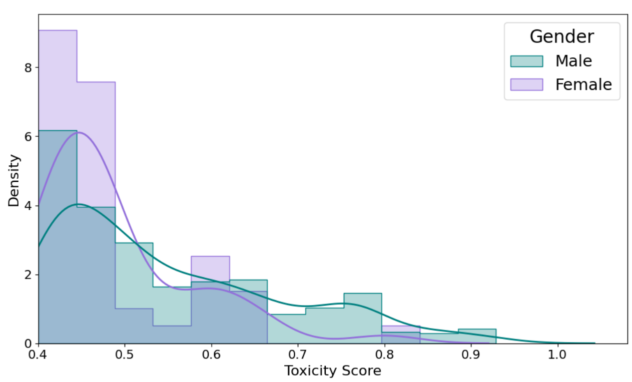

Figure 6: Density distribution of toxicity scores

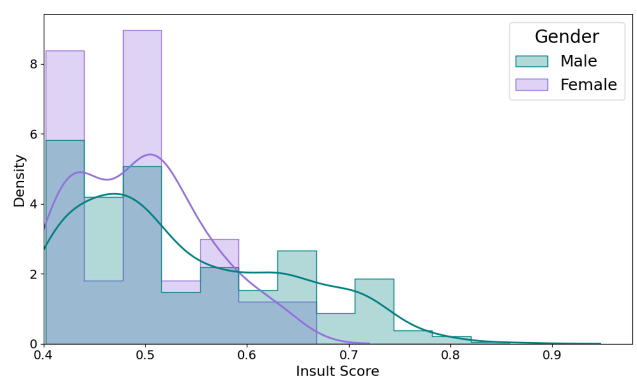

Figure 7: Density distribution of insult scores

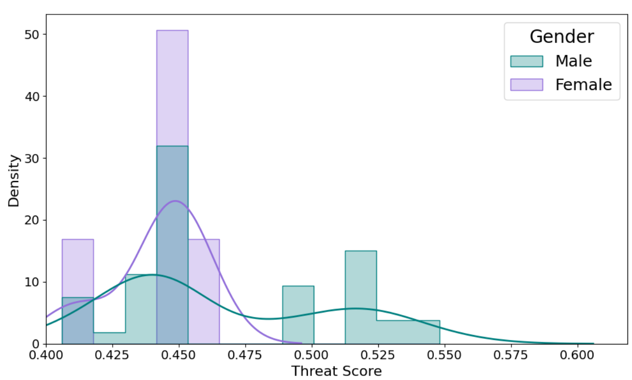

Figure 8: Density distribution of threat scores

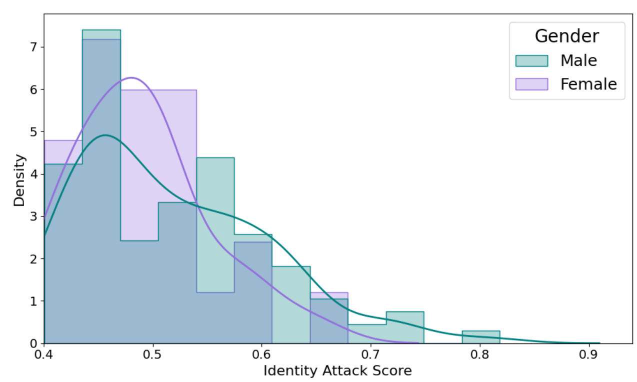

Figure 9: Density distribution of identity attack scores

3.7 Dataset inequality

Interestingly, only 28.7% of the gendered posts are made by female users, despite female users representing 33.56% of total posters in the dataset and 39% of Bluesky users [6]. This disparity reflects a gender gap that is often observed on social media platforms, where women tend to post less frequently [41]. However, further analysis reveals that much of this disparity is driven by outliers, including male posters with 100+ posts. As seen in Figures 10 and 11, the proportion of users with low post counts is quite similar across genders, while there are considerably more extreme high-volume male posters.

Figure 12 demonstrates that nearly 20% of all posts made by male users come from 9 users, each of whom made over 100 posts. Figure 13 illustrates that post distribution among male users is more unequal than among female users, though the overall inequality remains lower than that reported in previous studies on Twitter [54]. However, it is worth noting that these studies consider all posts made by users, whereas our analysis focuses solely on posts related to a specific topic.

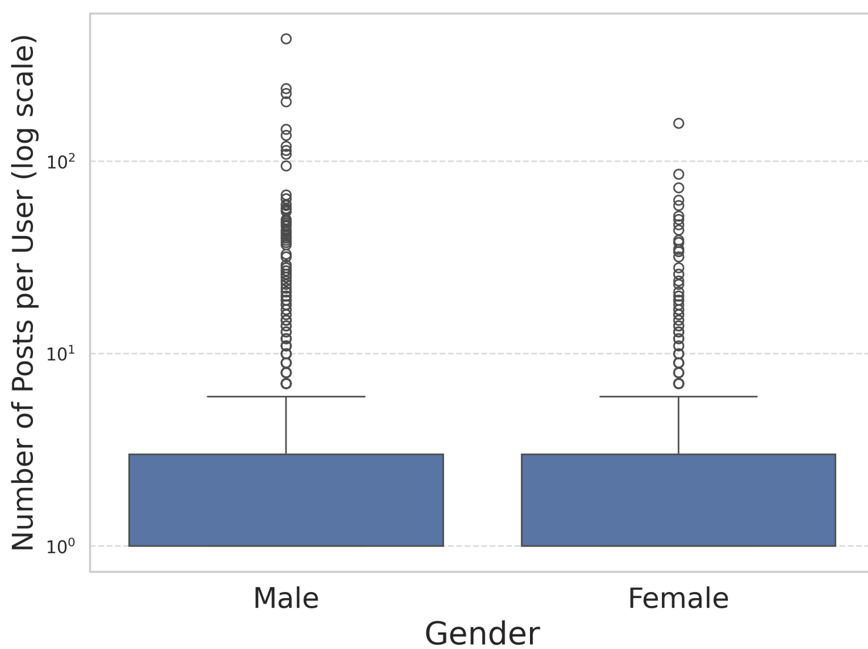

Figure 10: Logarithmic box-plot of posts by gender

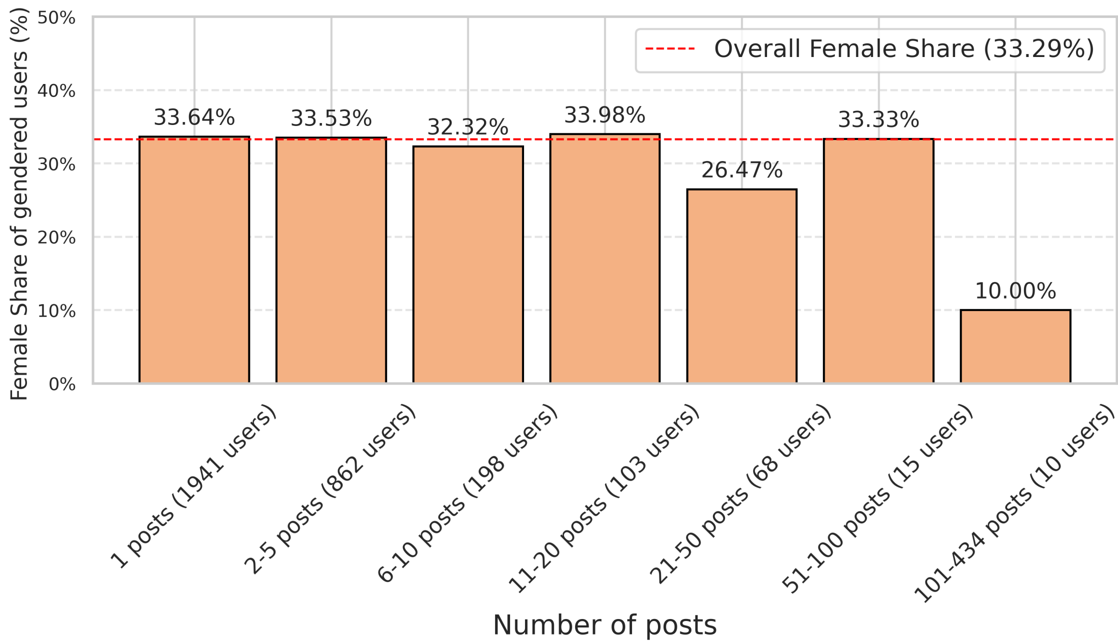

Figure 11: Proportion of male and female users who are female

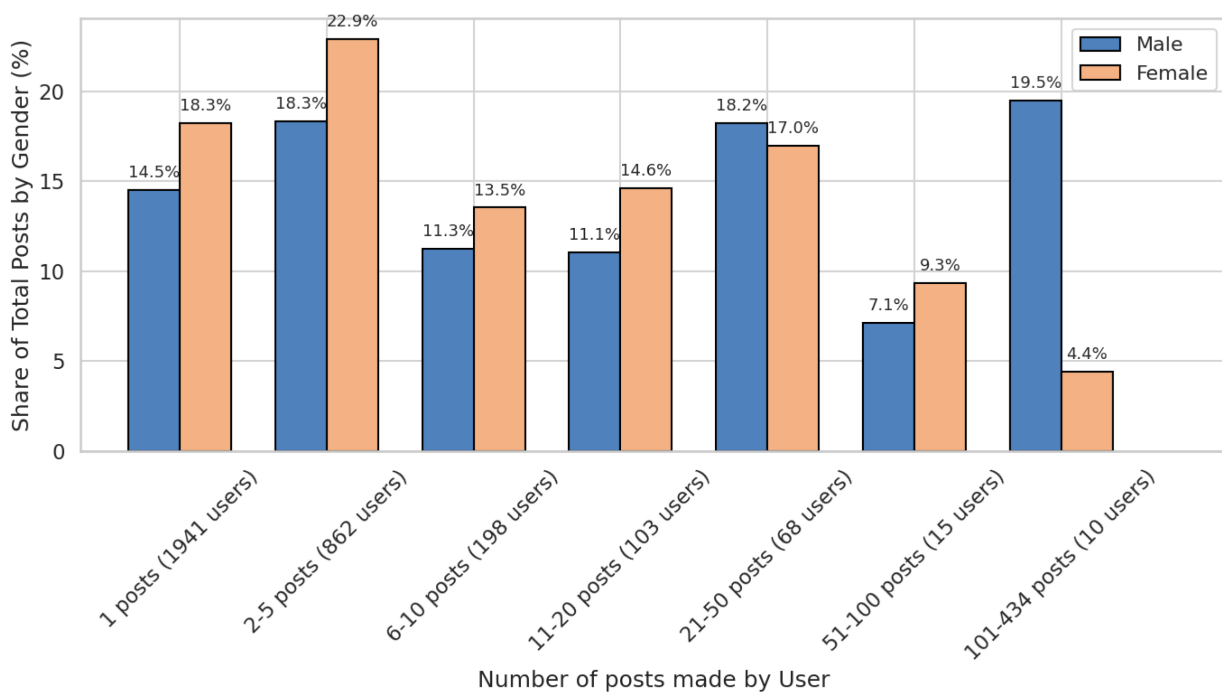

Figure 12: Distribution of posts by gender

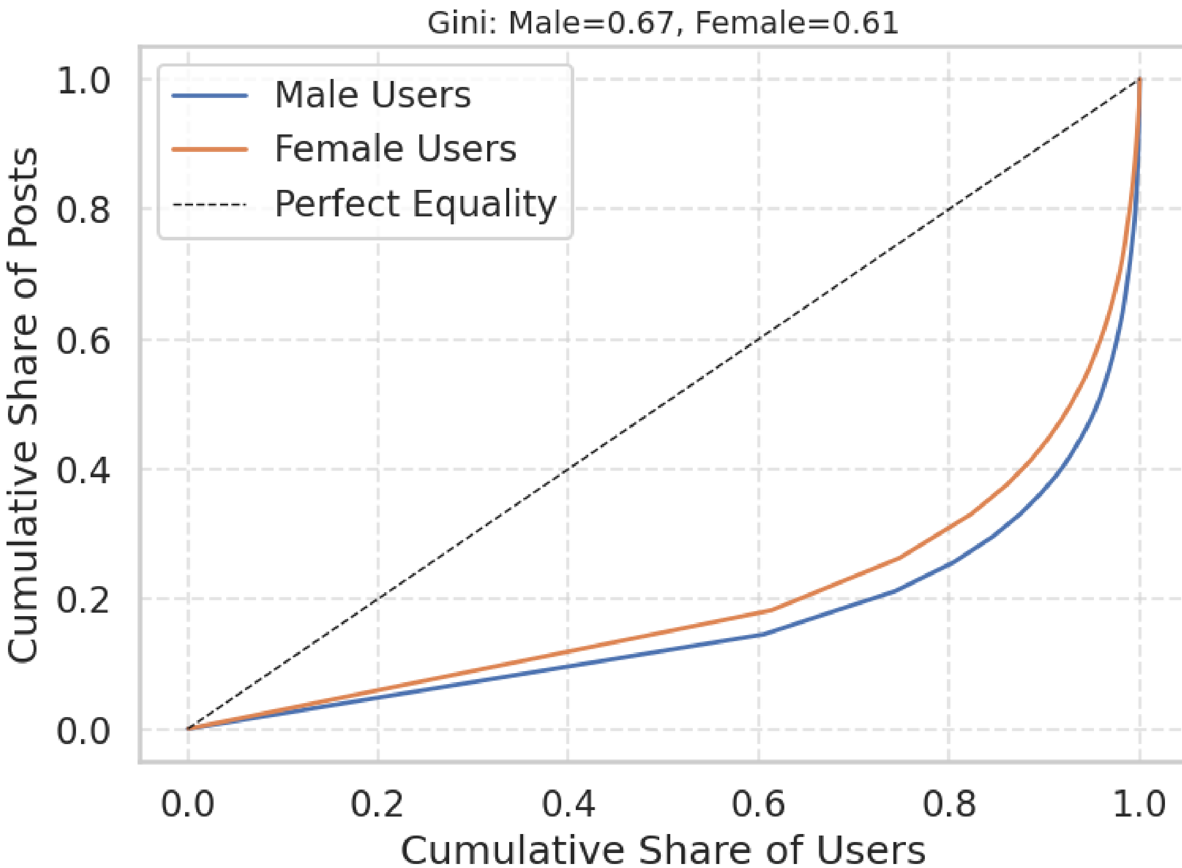

Figure 13: Lorenz curve of post distribution

3.8 Analysing Topics

To identify topics in the dataset, this study applied K-means clustering on text embeddings derived from the posts. Posts were cleaned by using the *tweet-preprocessor* library [7], which removed URLs, emojis, and numbers. Mentions were additionally removed using a regular expression, as the Bluesky format for mentions is not compatible with the library's mention handling. After preprocessing, TwHIN-BERT-base [52] was used to obtain 768-dimensional vector representations of the text. A random sample of 5,000 embeddings was first taken to calculate inertia scores for different values of k (number of clusters). The embeddings in the sample were normalized using the Euclidean (L2) norm. Afterwards, UMAP was applied to the normalized embeddings. The UMAP projection was then used to fit a K-means algorithm for each k between 4 and 13. The inertia for each k is shown in figure 14.

Based on the elbow method 14, the optimal value of k is 6, where the decrease in inertia starts to slow down. After finding the optimal k, we reapplied L2 normalization to all embeddings and ran K-means on the full set of embeddings.
After assigning the clusters, we used the NLTK Tweet-Tokenizer [3] to split all posts into words. We then removed stop words [1] and applied a lemmatizer [2] to extract word stems. For each cluster, the 30 most frequent words were counted. We then used a Chat-GPT (GPT-4.0) prompt (See appendix A) to generate a short, descriptive name of at most five words for each cluster. The resulting cluster names, size, and the 15 most frequent words per cluster are shown in Table 5

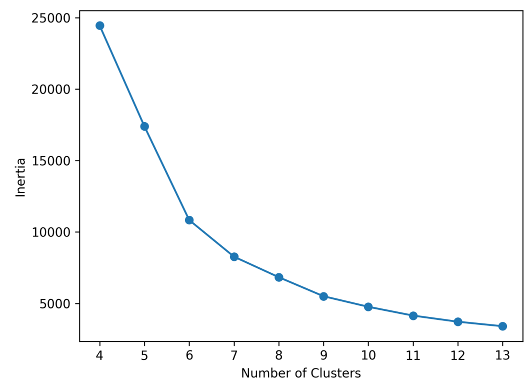

Figure 14: Elbow method: inertia vs number of clusters

Topic-ID	**Topic Name**	**Number of Tweets**	**Top 15 Most Frequent Words**
1	Dutch Parliament and Politics	10,591	tweedekamer, vvd, volledige, tekst, motie, pvv, wilder, nederland, kabinet, wel, bbb, mensen, gaat, aangenomen, weer
2	European Politics and Migration	1,956	europa, nederland, und, migranten, trump, ukraine, csu, photography, netherlands, cdu, fr, da, maastricht, eu, russland
3	Global Conflict and Politics	9,048	trump, ukraine, gaza, nato, israel, u, fascism, freepalestine, people, russia, fascist, eu, democrat, war, genocide
4	Right-Wing Dutch Politics	11,840	no, pvv, wilder, vvd, bbb, via, nsc, schoof, nederland, kabinet, faber, wel, weer, gaza, trump
5	News and Political Events	4,603	nieuws, rtl, no, arrestatie, npo, radio, usa, trump, denhaag, kabinet, navo, europa, video, gaza, verkiezingen
6	Dutch Jobs and Government Services	785	maastricht, vacature, bekijk, randstad, ministerie, utrecht, defensie, denhaag, hague, dordrecht, medewerker, adecco, gemeente, nederland, umc

Table 5: Assigned cluster names, post counts, and top 15 most frequent words per cluster

Figure 15: **size of topics**

Topic-ID	**Mean Sentiment**	**Mean Post Length**
1	0.292 (± 0.861)	252 (± 46)
2	0.270 (± 0.544)	192 (± 85)
3	-0.125 (± 0.767)	172 (± 97)
4	0.376 (± 0.819)	118 (± 55)
5	0.954 (± 0.222)	72 (± 29)
6	0.997 (± 0.050)	222 (± 34)
Overall mean	0.321 (± 0.818)	167.5 (± 90)

Table 6: **Mean post sentiment & length by topic**

In Figure 16, we observe that three of our clusters are composed almost entirely of neutral users. Examining the data within these clusters helps explain this pattern. Topic 5, *News and Political Events*, consists predominantly of posts from a single account that shares links to news articles in a consistent format, the title of the article with hashtags, explaining the short post length (6). This account is responsible for 92.4% of the posts in this topic. The posts in this topic are extremely positive, with a mean sentiment of 0.954, as shown in Table 6.

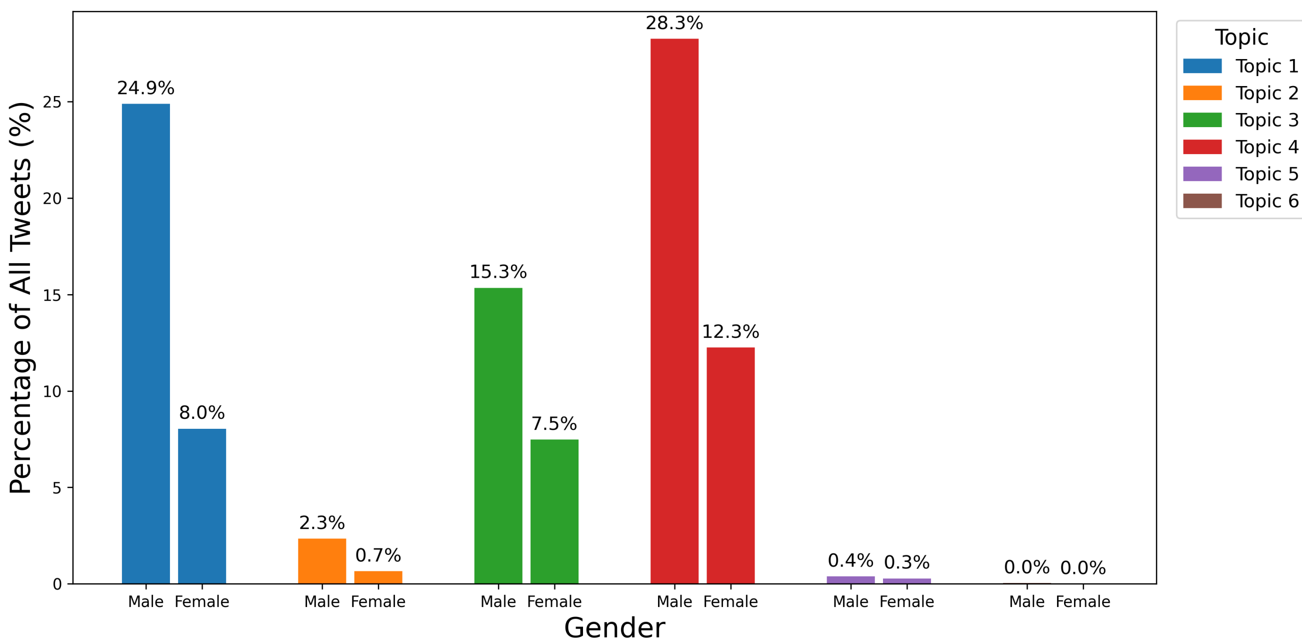

Figure 16: topics vs gender

Topic 6, *Dutch Jobs and Government Services*, is almost exclusively made up of job vacancy posts, again dominated by a single account making 50.4% of the posts. This is the cluster with the most positive sentiment, with a mean sentiment score of 0.997. Topic 2, titled *European Politics and Migration*, contains many posts from news accounts, like Topic 5. However, it is more diverse when it comes to the users of the topic. The top contributor in this cluster, also a news account, is responsible for 26.7% of the posts. As shown in Figure 17, this cluster has a relatively large number of English posters.

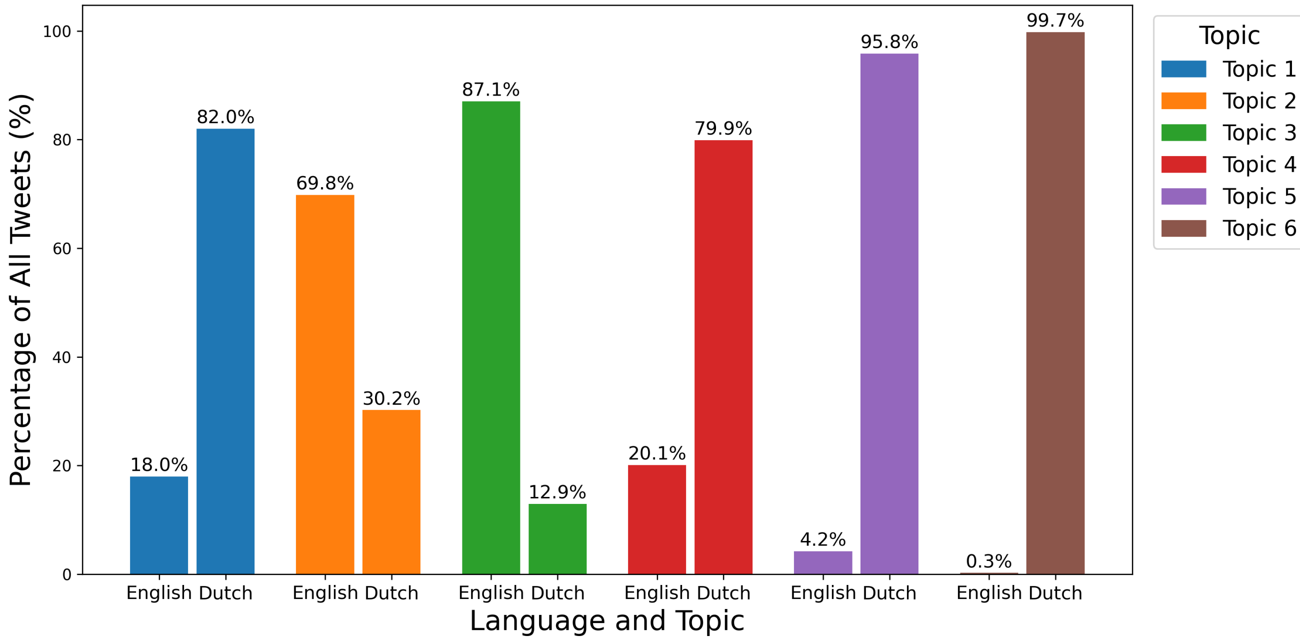

Figure 17: Topics vs language

The remaining three topics exhibit greater diversity in both content and user participation. Cluster 1, named *Dutch Parliament and Politics*, is more diverse compared to the previously discussed clusters when it comes to user participation, though the most prevalent account still contributes 11% of the posts. The cluster covers a broad range of topics related to Dutch policy and political decisions. It boasts the longest posts and average sentiment scores. This cluster has a relatively higher amount of male users compared to the other two clusters, with a lot of gendered users (Figure 18), suggesting male users tend to create more posts related to Dutch policy.

Cluster 3, *Global Conflict and Politics*, is the least centralized around a single user; its most active poster accounts for only 1.5% of the posts. The posts mostly discuss international politics and are primarily in English, as shown in Figure 17. This cluster has the most negative mean sentiment, being the only cluster that has a negative mean sentiment score.

Cluster 4, *Right-Wing Dutch Politics*, shows less centralization around a single user, with its top user contributing 6% of posts. The content is largely focused on criticisms of the Dutch government.

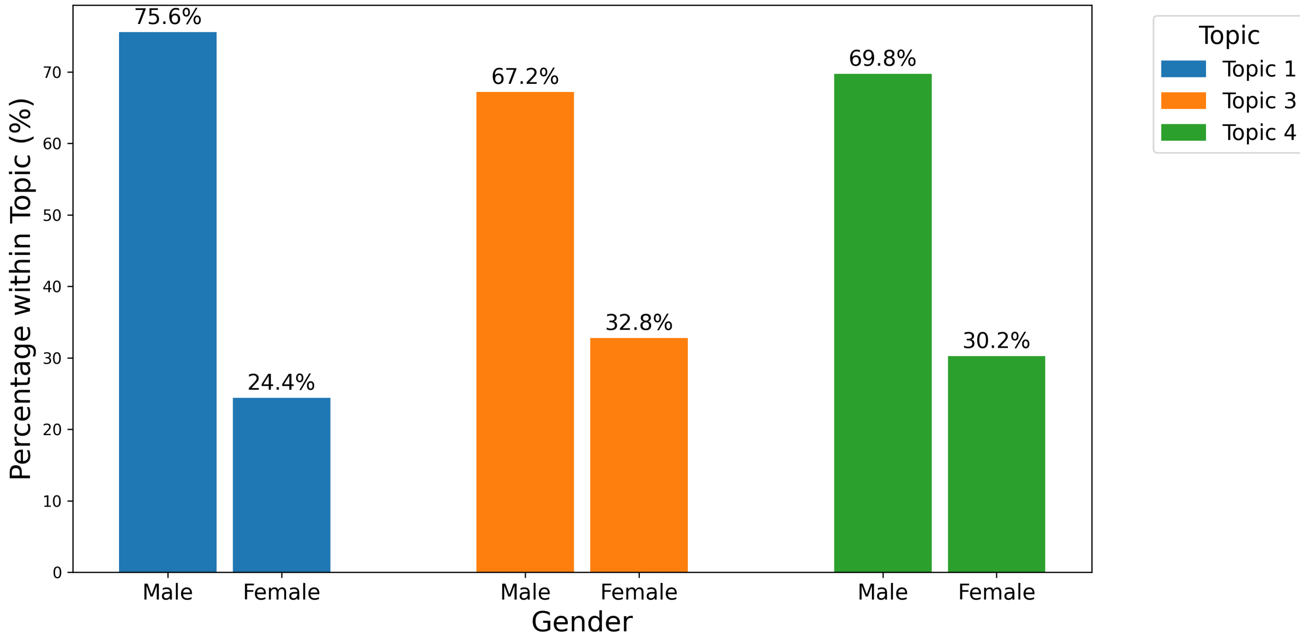

Figure 18: Normalized percentage of tweets per topic by gender (topics 1, 3, and 4).

Topic-ID	**Toxicity**	**Severe Toxicity**	**Insult**	**Threat**	**Identity attack**
1	0.167 ($\pm$ 0.166)	0.067 ($\pm$ 0.130)	0.143 ($\pm$ 0.178)	0.016 ($\pm$ 0.041)	0.080 ($\pm$ 0.130)
2	0.120 ($\pm$ 0.143)	0.037 ($\pm$ 0.109)	0.075 ($\pm$ 0.130)	0.018 ($\pm$ 0.045)	0.055 ($\pm$ 0.093)
3	0.202 ($\pm$ 0.201)	0.048 ($\pm$ 0.116)	0.119 ($\pm$ 0.173)	0.056 ($\pm$ 0.097)	0.125 ($\pm$ 0.162)
4	0.148 ($\pm$ 0.156)	0.050 ($\pm$ 0.112)	0.125 ($\pm$ 0.167)	0.017 ($\pm$ 0.049)	0.057 ($\pm$ 0.107)
5	0.049 ($\pm$ 0.062)	0.014 ($\pm$ 0.038)	0.022 ($\pm$ 0.033)	0.010 ($\pm$ 0.021)	0.020 ($\pm$ 0.041)
6	0.031 ($\pm$ 0.022)	0.011 ($\pm$ 0.007)	0.011 ($\pm$ 0.004)	0.006 ($\pm$ 0.000)	0.003 ($\pm$ 0.002)

Table 7: Mean Perspective scores by topic

From the perspective scores by topic (Table 7), we observe that Topics 5 and 6 have the lowest *toxicity*, which aligns with their more neutral content. The more diverse clusters (Topics 1,3, and 4) exhibit higher levels of *insult*. Topic 3 notably also shows relatively high scores for both *threat* and *identity attack*, but this could also be influenced by the perspective API detecting English-language content more reliably, as this Topic has a relatively large share of English posts.

Topic	Joy	Anger	Disgust	Fear	Sadness	Surprise	Trust	Anticipation
1	0.29 ± 0.19	0.10 ± 0.16	0.11 ± 0.12	0.02 ± 0.05	0.06 ± 0.06	0.03 ± 0.02	0.02 ± 0.01	0.25 ± 0.12
2	0.33 ± 0.31	0.13 ± 0.19	0.14 ± 0.15	0.02 ± 0.05	0.05 ± 0.07	0.03 ± 0.02	0.01 ± 0.01	0.21 ± 0.16
3	0.19 ± 0.30	0.28 ± 0.28	0.19 ± 0.16	0.05 ± 0.15	0.05 ± 0.13	0.01 ± 0.02	0.01 ± 0.01	0.12 ± 0.18
4	0.26 ± 0.21	0.12 ± 0.16	0.13 ± 0.12	0.02 ± 0.05	0.09 ± 0.09	0.03 ± 0.02	0.01 ± 0.01	0.23 ± 0.12
5	0.21 ± 0.15	0.11 ± 0.14	0.15 ± 0.12	0.03 ± 0.07	0.10 ± 0.07	0.04 ± 0.02	0.01 ± 0.01	0.26 ± 0.12
6	0.32 ± 0.13	0.05 ± 0.06	0.07 ± 0.04	0.01 ± 0.00	0.06 ± 0.03	0.04 ± 0.01	0.02 ± 0.00	0.33 ± 0.09

Table 8: Mean emotion scores per topic

In Table 8, the emotion scores are shown per topic. Topics 5 and 6, which also had the most positive mean sentiment, show high anticipation and *joy*, while displaying low *anger* and *sadness*. Topic 3, the most negative cluster, shows the highest *anger*, *disgust*, and *fear*, and the lowest surprise and anticipation. Topics 2 and 4 both show mixed emotional patterns, with moderately high disgust and sadness. Finally, Topic 1 shows moderate *joy* and *anticipation*, along with low levels of negative emotions.

Overall, our analysis reveals a clear relationship between topic composition and the emotional and toxic tone of posts. The clusters dominated by posts with set formats, Topics 5 and 6, tend to exhibit highly positive sentiment, low toxicity, and positive emotions. In contrast, more diverse clusters, particularly topics 3 and 4, are characterized by higher levels of negative emotions and toxic language, reflecting more critical discourse. These patterns underscore the importance of both user diversity and language in shaping the tone and emotional profile of online political discussions.

4 Discussion

This study provides a gender-based analysis of political discourse on the decentralized social media platform Bluesky, based on differences in sentiment, emotion, toxicity, and topical engagement. The results show several meaningful patterns across these dimensions.

The majority of posts in the dataset were classified as having positive sentiment expressions, and the positive emotions joy and anticipation were the most prevalent emotions. This suggests that, despite the often polarizing nature of political discourse, much of the conversation on Bluesky has an optimistic or positive tone. A possible explanation would be that Bluesky's unique methods of moderation cut down on negativity.

There was a clear shift in sentiment, which may be related to the collapse of the Dutch government, which happened during a similar period as the switch in sentiment. However, we can not prove causality and the switch could be caused by other factors.

Posts in English were found to be more negative. This could be caused by both models having a different scale for negativity [12], the fact that most posts in English talk about the same topic (international politics), or other factors.

Gender differences were subtle, but female users expressed more positive and negative sentiment. The emotion analysis showed that women tend to exhibit more joy and sadness, while men express more anger, disgust, and anticipation, suggesting that men and women express both different positive and negative emotions. A regression showed that most negative emotions show correlation with negative sentiment and vice versa, with the exception of disgust showing a significant but small positive relation with sentiment, which could be due to a range of causes, perhaps due to the TweetNLP model not perfectly detecting emotions or due to noise. Since some of the emotions are rarely present in the dataset (like fear, surprise, and trust) it is hard to draw any type of conclusions about the differences in expression of the emotions between genders or their relation to sentiment.

We found that male users tend to post more on average. This would be consistent with previous findings [41]. However, a big part of this disparity is caused by a small group of male users who made more than a hundred posts, whilst the median post amount between genders was quite similar. The proportion of female users was similar to that of the overall Bluesky population, suggesting that discourse about Dutch politics does not exhibit a specific participation gap. However, a general gender participation gap remains present on Bluesky. When looking at the overall distribution of post's per user, we find that while there is slightly more inequality between male posters, the overall inequality is still quite low when compared to previous research done on twitter [54].

In addition to sentiment and emotion, we examined the presence of toxic language. While overall toxicity levels were low, posts authored by male users were consistently more toxic than those written by female users. This aligns with previous research [36] and could be caused by social norms surrounding masculinity, differences in online communication styles between genders, or other factors.

Our topic modelling reveals some interesting patterns. Tweets about international politics were more likely to be in English and were more negative on average. Some clusters contain tweets with extremely consistent formatting, language, and sentiment. The more diverse clusters show similar patterns as well as clear differences. The topics suggest that male users are more likely to discuss policies and parliamentary motions, while female users are more likely to talk about international politics and conflicts. The topic discussing international politics is mostly in English, and is the most negative cluster with quite a big margin. The biggest cluster, 'Right-Wing Dutch Politics', has the shortest posts out of all the clusters, and the most positive out of the non-automated clusters.

In summary, this study provides a comprehensive gender-based analysis on political discourse on Bluesky, highlighting subtle but relevant differences in sentiment, emotion, toxicity, and topical engagement. Overall, posts in the dataset tend to be positive, and positive emotions are the most prevalent. Female users showed a slightly higher amount of positive sentiment, whilst both genders expressed different emotions at different rates. Male users post at a higher rate than female users, but part of this disparity is caused by a small subset of very high volume posters. Topic analysis found different topics showed different sentiment and emotion patterns, as well as having different poster demographics. Together, these findings illustrate that there are valuable patterns to be found in the way users communicate on Bluesky

4.1 Limitations

While this thesis provides valuable insights, several limitations should be considered when interpreting the findings. For this research, we ran the *search_posts()* method from the ATProto API [4] every day. This is not guaranteed to collect all posts, and posts in more active hashtags are more likely not to be collected, introducing extra noise into the dataset. For further research, we recommend using the Firehose API [13] instead, as it will not miss any posts.

Whilst collecting posts based on relevant hashtags gathers plenty of relevant posts, it can also collect plenty of posts that are not as relevant. It can also miss some relevant posts if they do not contain hashtags. The amount of data collected is also not extremely large, so a longer period of data collection would be recommended for better results.

Using two different sentiment models means the sentiment scores of the two languages are not guaranteed to be very comparable (as demonstrated by Bal et al. [12]), as one of the models might give some sentiment scores at a higher rate. The language detection from Bluesky's metadata is also not without its flaws, as some tweets in the dataset are German tweets identified as Dutch, or posts mostly in a different language that contain some English words.

NLP models are known to have inherent biases. These biases, like female expressed sentiment being detected more [49] or sentiment changing when the text is about male versus female subjects, or between European and African names, could influence the results of this study.

Although gender inference using an API is a good method, it is not guaranteed to predict genders completely correctly, introducing some extra noise to the dataset. The API is also not able to infer a gender for users who do not use their real name as their display name, losing valuable data on a big subset of the user base. Another downside to our method is that it only predicts user to be male or female, and thus does not account for other gender identities users might identify themselves with

Even though the emotion analysis and perspective analysis models are both multilingual, their performance is significantly better on English content. For optimal performance, it would be better to fine-tune a model on Dutch data specifically.

While using K-means clustering on TwHIN-BERT embeddings offers an efficient approach to topic modelling, this method has several limitations. Firstly, unsupervised clustering on noisy data means the resulting clusters are not guaranteed to be meaningful or human-interpretable topics. Secondly, dimensionality reduction using UMAP can distort the original embedding space. Finally, the pre-processing steps, such as removing mentions, URLs, and emojis, which are necessary for model compatibility, may eliminate meaningful context.

5 Conclusions

This thesis offers an in-depth exploration of how sentiment, emotion, and toxicity are expressed on Bluesky. Through the application using state-of-the-art NLP tools and clustering methods, the study found that emotional and sentimental expression varies significantly across genders, languages, and topical engagement. Female users expressed more joy and sadness alongside slightly more negative sentiment, whereas male users more frequently expressed anger, disgust, and anticipation. Toxicity levels were generally low, though higher in male-authored posts. Topic modelling revealed six distinct clusters, reflecting key areas of Dutch political discussion, each exhibiting different patterns of sentiment and user diversity. Some clusters were dominated by one or a few users, whilst others were more diverse. Each cluster showed differences in sentiment and emotion expression, language and gender distribution, and topical content.

These findings directly address the central research question, showing that Bluesky is shaped by user demographics, language, and topical focus. Although the platform's decentralized nature and relatively small, specific user base differentiate it from mainstream social media, many familiar patterns in communication persist.

5.1 Further Research

The field of sentiment analysis on Bluesky is still relatively young and unexplored, leaving plenty of research opportunities to be explored.

Further research could look at interactions between users, including comments, likes, and reposts in the dataset. This could reveal patterns such as differences in cross-gender versus intra-gender communication. These could also be used to explore the network structure of Dutch political Bluesky, investigating who influences whom and what communities have formed.

Further research could also develop specific Dutch models for emotion and perspective analysis by fine-tuning existing models to Dutch data. Additionally, a study could look at political discourse across multiple countries and languages, and potentially reveal more nation-specific patterns. Furthermore, as Bluesky continues to grow, longer studies could explore how the platform evolves.

Another research opportunity would be comparative platform studies, where Bluesky would be compared to other platforms like X or Reddit. Comparative studies could also be done between countries and languages, to see how language, political and cultural differences influence online sentiment and emotion expression.

Future research can build on this thesis to improve our understanding of political expression and social dynamics in emerging digital spaces.

References

[1] Dutch stopwords list from node-nltk-stopwords. https://github.com/xiamx/node-nltk-stopwords/blob/master/data/stopwords/dutch. Accessed: 2025-07-22.

[2] nltk.stem.wordnetlemmatizer — nltk 3.8.1 documentation. https://www.nltk.org/api/nltk.stem.WordNetLemmatizer.html?highlight=wordnet. Accessed: 2025-07-22.

[3] nltk.tokenize.casual — nltk 3.8.1 documentation. https://www.nltk.org/api/nltk.tokenize.casual.html. Accessed: 2025-07-22.

[4] At protocol api documentation. https://docs.bsky.app/, 2025. Accessed: 2025-07-17.

[5] Nos.nl - nieuws, sport en evenementen. https://nos.nl/collectie/13996, 2025. Accessed: 2025-07-17.

[6] Similarweb - website analytics for bsky.app. https://www.similarweb.com/website/bsky.app/, 2025. Accessed: 2025-07-17.

[7] Tweet preprocessor. https://pypi.org/project/tweet-preprocessor/, 2025. Python package on PyPI. Accessed: 2025-07-17.

[8] R. B. Andeweg, Galen A. Irwin, and Tom Louwerse. *Governance and politics of the Netherlands*. Bloomsbury Academic, 2020.

[9] Valentina Andries and Ailsa Niven. Social media research: Ethical guidance for researchers at the university of edinburgh. 2021.

[10] A. arn, G. Achim, and E. Dubencu. White paper: Name genderizing.

[11] Michele Avalle, Niccolò Di Marco, Gabriele Etta, Emanuele Sangiorgio, Shayan Alipour, Anita Bonetti, Lorenzo Alvisi, Antonio Scala, Andrea Baronchelli, Matteo Cinelli, et al. Persistent interaction patterns across social media platforms and over time. *Nature*, 628(8008):582–589, 2024.

[12] DM Bal, MA Bal, A {v. Bunningen}, Alexander Hogenboom, Frederik Hogenboom, and Flavius Frasincar. Sentiment analysis with a multilingual pipeline. In A. Bouguettaya, M. Hauswirth, and L. Liu, editors, *Web Information System Engineering*, volume 6997, pages 129–142, Germany, October 2011. Springer-Verlag.

[13] Bluesky. Firehose - advanced guides, 2024. Accessed: 2025-07-25.

[14] Jose Camacho-Collados, Kiamehr Rezaee, Talayeh Riahi, Asahi Ushio, Daniel Loureiro, Dimosthenis Antypas, Joanne Boisson, Luis Espinosa-Anke, Fangyu Liu, Eugenio Martínez-Cámara, et al. TweetNLP: Cutting-Edge Natural Language Processing for Social Media. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Abu Dhabi, U.A.E., November 2022. Association for Computational Linguistics.

[15] Charles Crabtree, Matt Golder, Thomas Gschwend, and Indrii H Indriason. It is not only what you say, it is also how you say it: The strategic use of campaign sentiment. *The Journal of Politics*, 82(3):1044–1060, 2020.

[16] Luna De Bruyne, Orphée De Clercq, and Véronique Hoste. Emotional robbert and insensitive bertje: combining transformers and affect lexica for dutch emotion detection. In *Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), held in conjunction with EACL 2021*, pages 257–263. Association for Computational Linguistics, 2021.

[17] Wietse De Vries, Andreas van Cranenburgh, Arianna Bisazza, Tommaso Caselli, Gertjan van Noord, and Malvina Nissim. Bertje: A dutch bert model. *arXiv preprint arXiv:1912.09582*, 2019.

[18] Pieter Delobelle, Thomas Winters, and Bettina Berendt. Robbert: a dutch roberta-based language model. *arXiv preprint arXiv:2001.06286*, 2020.

[19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

[20] Alejandro Garcia-Rudolph, Sara Laxe, Joan Saurí, and Montserrat Bernabeu Guitart. Stroke survivors on twitter: sentiment and topic analysis from a gender perspective. *Journal of medical Internet research*, 21(8):e14077, 2019.

[21] Erwin Gielens, Femke Roosma, and Peter Achterberg. Between left and right: A discourse network analysis of universal basic income on dutch twitter. *Journal of Social Policy*, page 1–22, 2023.

[22] Frederic Guerrero-Solé and Cristina Perales-García. Bridging the gap: How gender influences spanish politicians' activity on twitter. *Journalism and media*, 2(3):469–483, 2021.

[23] Tzung-Pei Hong, Leticia Serrano-Estrada, Akrati Saxena, and Anupam Biswas. *Deep learning for social media data analytics*. Springer, 2022.

[24] Andreas M. Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, 53(1):59–68, 2010.

[25] Svetlana Kiritchenko and Saif M Mohammad. Examining gender and race bias in two hundred sentiment analysis systems. *arXiv preprint arXiv:1805.04508*, 2018.

[26] Funda Kivran-Swaine, Sam Brody, Nicholas Diakopoulos, and Mor Naaman. Of joy and gender: emotional expression in online social networks. In *Proceedings of the ACM 2012 conference on computer supported cooperative work companion*, pages 139–142, 2012.

[27] Martin Kleppmann, Paul Frazee, Jake Gold, Jay Graber, Daniel Holmgren, Devin Ivy, Jeromy Johnson, Bryan Newbold, and Jaz Volpert. Bluesky and the at protocol: Usable decentralized social media. In *Proceedings of the ACM Conext-2024 Workshop on the Decentralization of the Internet*, CoNEXT '24, page 1–7. ACM, December 2024.

[28] S Kosmidis, S Hobolt, E Molloy, and S Whitefield. Party polarization and emotive rhetoric. *Comparative Political Studies, forthcoming*, 2018.

[29] Sudhanshu Kumar, Monika Gahalawat, Partha Pratim Roy, Debi Prosad Dogra, and Byung-Gyu Kim. Exploring impact of age and gender on sentiment analysis using machine learning. *Electronics*, 9(2), 2020.

[30] Alyssa Lees, Vinh Q Tran, Yi Tay, Jeffrey Sorensen, Jai Gupta, Donald Metzler, and Lucy Vasserman. A new generation of perspective api: Efficient multilingual character-level transformers. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 3197–3207, 2022.

[31] Mariana Macedo and Akrati Saxena. Gender differences in online communication: A case study of soccer. *arXiv preprint arXiv:2403.11051*, 2024.

[32] Cristina Manzano and Juan A Sánchez-Giménez. Women, gender and think tanks: political influence network in twitter 2018. 2019.

[33] Benjamin Muller, Antonis Anastasopoulos, Benoît Sagot, and Djamé Seddah. When being unseen from mbert is just the beginning: Handling new languages with multilingual language models. *arXiv preprint arXiv:2010.12858*, 2020.

[34] Gianluca Nogara, Erfan Samieyan Sahneh, Matthew R. DeVerna, Nick Liu, Luca Luceri, Filippo Menczer, Francesco Pierri, and Silvia Giordano. A longitudinal analysis of misinformation, polarization and toxicity on bluesky after its public launch, 2025.

[35] Delfín Ortega-Sánchez, Joan Pagès Blanch, Jaime Ibáñez Quintana, Esther Sanz de la Cal, and Raquel de la Fuente-Anuncibay. Hate speech, emotions, and gender identities: A study of social narratives on twitter with trainee teachers. *International journal of environmental research and public health*, 18(8):4055, 2021.

[36] Rrubaa Panchendrarajan, Harsh Saxena, and Akrati Saxena. Social media and academia: How gender influences online scholarly discourse. *arXiv preprint arXiv:2505.03773*, 2025.

[37] Róisín Parkins. *Gender and emotional expressiveness: An analysis of prosodic features in emotional expression*. Griffith University Nathan, QLD, 2012.

[38] Md Mostafizer Rahman, Ariful Islam Shiplu, Yutaka Watanobe, and Md Ashad Alam. Roberta-bilstm: A context-aware hybrid model for sentiment analysis. *IEEE Transactions on Emerging Topics in Computational Intelligence*, pages 1–18, 2025.

[39] Vikas Reddy and Giovanni Luca Ciampaglia. Murkysky: Analyzing news reliability on bluesky, 2025.

[40] Ludovic Rheault, Kaspar Beelen, Christopher Cochrane, and Graeme Hirst. Measuring emotion in parliamentary debates with automated textual analysis. *PloS one*, 11(12):e0168843, 2016.

[41] Chiara Rollero, Adriano Daniele, Stefano Tartaglia, et al. Do men post and women view? the role of gender, personality and emotions in online social activity. *Cyberpsychology*, 13(1):1–15, 2019.

[42] Ali Salloum, Dorian Quelle, Letizia Iannucci, Alexandre Bovet, and Mikko Kivelä. Politics and polarization on bluesky, 2025.

[43] Akrati Saxena, Harita Reddy, and Pratishtha Saxena. Introduction to sentiment analysis covering basics, tools, evaluation metrics, challenges, and applications. *Principles of social networking: the new horizon and emerging challenges*, pages 249–277, 2021.

[44] Akrati Saxena, Harita Reddy, and Pratishtha Saxena. Recent developments in sentiment analysis on social networks: techniques, datasets, and open issues. *Principles of Social Networking: The New Horizon and Emerging Challenges*, pages 279–306, 2021.

[45] Gijs Schumacher, Daniel Hansen, Mariken A.C.G. van der Velden, and Sander Kunst. A new dataset of dutch and danish party congress speeches. *Research & Politics*, 6(2):2053168019838352, 2019.

[46] Wladimir Sidorenko. Sentiment analysis of german twitter. *CoRR*, abs/1911.13062, 2019.

[47] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. How to fine-tune bert for text classification? In *China national conference on Chinese computational linguistics*, pages 194–206. Springer, 2019.

[48] Nirmalya Thakur, Shuqi Cui, Karam Khanna, Victoria Knieling, Yuvraj Nihal Duggal, and Mingchen Shao. Investigation of the gender-specific discourse about online learning during covid-19 on twitter using sentiment analysis, subjectivity analysis, and toxicity analysis. *Computers*, 12(11):221, 2023.

[49] Mike Thelwall. Gender bias in sentiment analysis. *Online Information Review*, 42(1):45–57, 2018.

[50] Sophie F Waterloo, Susanne E Baumgartner, Jochen Peter, and Patti M Valkenburg. Norms of online expressions of emotion: Comparing facebook, twitter, instagram, and whatsapp. *New media & society*, 20(5):1813–1831, 2018.

[51] Benedict Witzenberger and Jürgen Pfeffer. More inclusive and wider sources: A comparative analysis of data and political journalists on twitter (now x) in germany. *Journalism and Media*, 5(1):412–431, 2024.

[52] Xinyang Zhang, Yury Malkov, Omar Florez, Serim Park, Brian McWilliams, Jiawei Han, and Ahmed El-Kishky. Twhin-bert: A socially-enriched pre-trained language model for multilingual tweet representations. *arXiv preprint arXiv:2209.07562*, 2022.

[53] Yixuan Zhang, Bing Zhou, Yiyan Hu, and Kun Zhai. From individual expression to group polarization: A study on twitter's emotional diffusion patterns in the german election. *Behavioral Sciences*, 15(3):360, 2025.

[54] Linhong Zhu and Kristina Lerman. Attention inequality in social media. *arXiv preprint arXiv:1601.07200*, 2016.

Appendix A: LLM Prompt for Naming Clusters

```
Prompt:

Below are the top 30 words for each cluster generated by a K-means clustering
↪ model applied to tweets.
Please analyse the words in each cluster and assign a concise name (maximum 5
↪ words) that best summarizes the overall theme or topic of the cluster.

Return the result in the following format:

Cluster 0: [Name]
Cluster 1: [Name]
...

Cluster Data:

Cluster 0:
['tweedekamer', 'vvd'..., ...]

Cluster 1:
['europa', 'nederlands, etc.]
```