

The Linguistic Landscape of “Controversial”: Sentiment and Theme Distribution Insights

Elizaveta G. Grishechko
grishechko-eg@rudn.ru
RUDN University, Russia

ABSTRACT

The language used to frame controversial topics on social media has profound implications for public discourse and opinion formation, warranting a close examination of their sentiment and thematic distribution. This study investigates the sentiment and themes associated with controversial topics by analyzing Reddit posts containing the token “controversial” in their titles on three news-related subreddits, aiming to bridge a gap in existing literature by focusing on platform-specific sentiment analysis with an emphasis on content typology. A mixed-methods NLP approach instrumented via Python was employed, combining VADER-supported sentiment analysis and a qualitative content analysis using n-grams to identify and categorize themes. The sentiment analysis results indicated that most of the content had neutral sentiment, which testifies to the predominantly fact-based approach to presenting information with lack of strong emotional connotations. However, the overall compound sentiment scores were negative, which suggests a strong negative undertone in the framing of controversial topics. The theme distribution analysis revealed that Politics and Legislation was the most predominant theme, followed by Technology and Surveillance, Social Issues and Controversies, Health and Medicine, and Environment and Energy. This distribution attests to a range of societal issues that generate controversy on social media platforms. Study findings can be used by content creators and social media analysts to track online content sentiment, guide content moderation practices, and improve audience engagement. By demonstrating the potential of NLP techniques, this study also contributes to the fields of media research and language technology, which can encourage a better scholarly evaluation of online discourse.

Keywords: sentiment analysis; natural language processing; social media; online discourse; Python

INTRODUCTION

The rapid proliferation of social media platforms has significantly impacted the manner in which news is disseminated, consumed, and discussed. Numerous studies have pointed out the impact of digital discourse on public opinion, emphasizing the need for a scholarly delineation of the linguistic strategies employed in controversial discussions (Wu & Pan, 2022; Splendiani & Capriello, 2022; Aziz & Hashim, 2021). The language used to present news stories naturally plays a crucial role in shaping public opinion and discourse, and in this context sentiment analysis has emerged as an expedient tool to decode the affective states and subjective information behind linguistic choices.

Previous research on news media content focused on a variety of issues pertaining to measuring and interpreting news sentiment. For example, Shapiro et al. (2022) introduced a time-series measure of economic sentiment derived from newspaper articles to emphasize the utility of

combining existing lexicons and accounting for negation to enhance predictive accuracy. Their findings attested to the predictive value of news sentiment on consumer sentiment and demonstrated the impact of positive sentiment shocks on macroeconomic variables, such as consumption output and interest rates. Taj et al. (2019) described the capabilities of lexicon-based sentiment analysis of news articles in handling the vast amounts of data generated in the digital age. Their work, based on the BBC news dataset, validated the effectiveness of lexicon-based methods in discerning the emotions within news content, justifying its practical applicability in modern information technology environments. Kumar et al. (2018) investigated sentiment within social media news channels to expose how different news sources leverage sentiment to attract user engagement. Their analysis acknowledged a strong correlation between the sentiment of news posts and user reactions and outlined the varied strategies employed by digital television, radio, and print media. Oh et al. (2023) explored the differential effects of sentiment on digital news consumption and social media sharing to look at how negative and positive sentiments drive readership and sharing patterns, respectively. Their investigation detailed the contrasting effects of author sentiment and news topic valence on news engagement and offered guidance for news publishers on optimizing content for digital platforms. Waddell (2020) investigated the impact of comment sentiment on news evaluation through social media to show how negative comments could diminish news credibility and the perceived importance of issues. This study paid special attention to comment authenticity and sentiment and proved instrumental in describing audience interactions in digital news contexts.

This brief review of available research shows that sentiment analysis of news media content is well-trodden ground. But it also bears testimony to the prevalence of an all-encompassing approach that often yields generalized perceptions across aggregate data sources. We believe that such approaches, although valid in their own right, may overlook the specifics of news language and sentiment prevalent within different platforms and contexts, thus failing to capture the variations in news sentiment in different online environments. This, as we argue, signals the need for more narrowed approaches within platform-specific sentiment analysis studies. Additionally, we agree with Žitnik et al. (2022), who suggested that content typology should become the foundational principal in dataset collection, meaning that targeted differentiation between critical vs uncritical, time-sensitive vs non-time-sensitive, controversial vs uncontroversial content, etc. should be considered a guiding principle in content screening for research and interpretation purposes. Consequently, there's a clear gap in the literature regarding the exploration of platform-specific news sentiment dynamics with typological content differentiation.

To address this research gap, this study employs a combination of web scraping, natural language processing (NLP), and sentiment analysis techniques to investigate the sentiment of news articles posted on Reddit in three news-related subreddits and containing the token "controversial" in the title, and identify the key themes associated with its use in post titles. By adopting this focused lens, the study endeavors to offer a narrow examination of online news sentiment differentiated in terms of content typology and single-platform data sourcing.

Reddit was selected for data collection due to its fitting structural organization into subreddits, which are community-driven forums dedicated to specific topics. This structure facilitates the collection of topic-specific content, making it an ideal environment for analyzing discourse on controversial issues. Importantly, the choice is also justified by Reddit's scarce representation as a data source in the literature, let alone in the framework of controversial content analysis, with studies primarily focused on the already extensively discussed issues of public

engagement (Aldous et al., 2023), marketing (Machavarapu, 2022), opinion prediction (Huynh et al., 2021), or recent health crisis perceptions (Yan et al., 2021).

The token “controversial” is of particular interest in this study due to its unique semantics and application in social media discourse. The token inherently implies disagreement or conflict, which can be linked to the polarization and division present in many public discussions, especially on social media platforms (Kubin & von Sikorski, 2021). Building on this inherent semantics of the token, this study can screen controversial news posts effectively and expand on the nature of discourse surrounding contentious issues, including how these topics are framed in this particular online community.

To that end, the study will attempt to reflect on three research questions:

1. What is the predominant sentiment of linked news articles for Reddit posts containing the token “controversial” in their titles?
 2. What are the predominant themes associated with the use of “controversial” in news post titles on Reddit, and how are they distributed?
 3. What ensuing implications relevant to social media language research and practice can be deduced from the identified sentiment scores and theme distribution registered in Reddit posts?
- By addressing these research questions, the study can contribute to a better understanding of how controversial topics are framed on Reddit, specifically in news-related subreddits.

LITERATURE REVIEW

The rise of social media platforms has led to a surge in research on social media language, as scholars strive to understand how language evolves in these digital environments. Earlier studies focused on the unique linguistic features of social media discourse, such as the use of abbreviations, emojis, and unconventional grammar (Miyake, 2007; Kralj Novak, 2015). More recent research has expanded to examine the effects of social media language on various aspects of interaction, such as identity construction (Kasperuniene & Zydziunaite, 2019), persuasion (Agur & Frisch, 2019), information dissemination (Vosoughi et al., 2018; Agapova & Grishechko, 2016), and conflict resolution (Grishechko, 2023a).

Furthermore, research on social media language has looked into the role of linguistic precedence in shaping linguistic practices (Sibul et al., 2019; Malyuga & Akopova, 2021), as well as the impact of platform affordances, such as character limits and multimedia capabilities, on language use (Zappavigna, 2018). Discourse analysis has also emerged as a valuable approach to studying social media language, enabling researchers to explore the complex interconnection of language, power, and ideology in online interactions (Wu & Pan, 2022). Additionally, studies have investigated the influence of social media language on offline engagement, revealing that digital language practices can both reflect and shape broader linguistic trends (Lai & Fu, 2021).

Researchers have also looked into the role of social media language in the construction and negotiation of social relationships and group identities. For instance, Chau & Lee (2021) have explored how users employ linguistic strategies, such as code-switching, to navigate and signal belonging in diverse online communities, while Schoenebeck et al. (2023) addressed the challenges posed by online harassment and hate speech, pointing out the need for effective moderation tools and policies that account for linguistic nuances.

Moreover, the analysis of social media language in crisis communication (Splendiani & Capriello, 2022) and health-related discussions (Fernández-Luque & Bau, 2015) has proven

valuable in understanding public opinion, misinformation, and the role of language in shaping collective responses.

The rise of social media platforms has also led to an interest in researching memetic language and the factors that contribute to the virality of content. Memetic language refers to the use of memes, catchphrases, and other forms of easily shareable content that often combine text, images, and multimedia elements. This area of research explores how language is utilized in these digital environments to create relatable, humorous, or thought-provoking content that resonates with users and encourages sharing. In this field, linguistic research investigates the unique linguistic features of memes and other viral content, such as the use of humor, irony, intertextuality, and visual elements (Shifman, 2013).

Sentiment analysis has become a popular research area in social media language studies, as it allows for the examination of emotions, opinions, and attitudes expressed in digital texts. Researchers have employed a variety of computational techniques, including machine learning and natural language processing, to analyze sentiment in social media content (Kavitha et al., 2022; Wu et al., 2022; Omar & Al-Tashi, 2018). Applications of sentiment analysis in social media research range from understanding consumer preferences (Păvăloaia et al., 2019; Grishechko, 2023b) to monitoring public opinion on political events (Belcastro et al., 2022) and detecting mental health issues (Benrouba & Boudour, 2023). This exploration of sentiment analysis in social media language studies directly informs the methodology and objectives of our research. By leveraging computational techniques, this study applies natural language processing and sentiment analysis to dissect the sentiment surrounding news post titles on Reddit. Our work further expands upon the applications of sentiment analysis by focusing on controversial topics, aligning with the broader aims of understanding information framing in digital communication.

Thus, the growing body of research on social media language reflects the diverse ways in which language is employed, adapted, and transformed in digital contexts, unmasking the broader implications for society, culture, and discourse.

Given that this study is also concerned with the analysis of “controversial” within news post titles – a token characterized by its evaluative nature – it is also expedient to provide literature review on evaluative adjectives as part of media discourse.

Evaluative adjectives are descriptive words that express a judgment or assessment, often imbuing a statement with a positive or negative connotation regarding the qualities or characteristics of a noun (Bednarek, 2006). These adjectives play an important role in shaping public opinion and discourse in media contexts. They contribute to framing, a process in which certain aspects of a topic are emphasized to convey a particular interpretation or evaluation. Research on evaluative adjectives in media discourse has investigated their use in various contexts, such as political discourse (López-Rabadán, 2021), news reporting (Ash et al., 2019), and opinion articles (Johannessen, 2015). These studies have found that evaluative adjectives can influence readers’ perceptions and attitudes towards the topics being discussed, often by implicitly conveying the author’s stance or opinion. For example, in political discourse, adjectives like “repressive” or “progressive” can shape opinions about a government’s policies (López-Rabadán, 2021). In news reporting, terms such as “catastrophic” for a natural event or “innovative” for a technological advancement guide the audience’s emotional response and evaluation of the information presented (Ash et al., 2019). These examples illustrate how evaluative adjectives, by framing topics in a certain light, can affect public opinion and discourse, attesting to their importance in media and communication studies.

In addition to their impact on readers' perceptions, evaluative adjectives can also shape the overall tone and sentiment of media texts. According to Bednarek (2006), they are central to the expression of appraisal, which encompasses the linguistic resources used to convey judgments, evaluations, and emotions. Using evaluative adjectives, media outlets can subtly convey positive, negative, or neutral evaluations of events, people, or issues, thus potentially influencing public opinion and sentiment.

In the context of social media, evaluative adjectives take on added significance due to the highly interactive nature of online discourse, as well as their influential role in determining user engagement and virality (Berger & Milkman, 2012). In addition to framing, evaluative adjectives can serve various pragmatic functions, such as expressing solidarity with or opposition to a particular viewpoint (Grašič, 2022), or signaling group membership and identity (Mostafa, 2019). Research on evaluative adjectives in social media discourse remains limited. However, a few studies have examined their use in specific contexts, such as online comments (Zubir et al., 2021) and online reviews (Azhari & Fang, 2018). The selection of evaluative adjectives in social media discourse has been found to contribute to the spread of perspectives or stances, as well as the polarization of online discussions (Alkhamash, 2021). Furthermore, the use of evaluative adjectives has been proved to impact the credibility of news sources, as readers may perceive content framed with certain evaluative adjectives as more or less objective or biased (Tandoc et al., 2018). These studies suggest that evaluative adjectives can significantly impact the sentiment and tone of online discussions. For instance, a product review that describes an item as "innovative" can boost its appeal among users and promote positive engagement (Grishechko, 2023b). Conversely, labeling a political move as "disastrous" in an online comment can amplify opposition, influencing the tone and direction of the discourse.

This study specifically focuses on the evaluative adjective "controversial" to screen for news posts that are inherently deemed contentious by the authors in the Reddit community. This choice is not arbitrary; it is predicated on the understanding that the adjective "controversial" signals a divergence of opinions, marking the content as potentially polarizing or subject to debate. This screening mechanism allows us to capture a snapshot of the discourse surrounding topics that elicit strong reactions, making "controversial" not just a descriptor but a lens through which we analyze the sentiment and thematic distribution on Reddit. Following this, the study employs n-grams to further dissect and categorize themes within these posts, showing how controversial topics are framed across different subreddits. This methodological approach underscores the token's role not just in identifying relevant content but also in structuring the thematic analysis.

METHODOLOGY

DATA COLLECTION

The methodology employed in this study involves a combination of web scraping, natural language processing (NLP), and sentiment analysis techniques to collect and analyze data from news-related subreddits. NLP combines computational techniques with models of language understanding to perform tasks such as translating texts from one language to another, responding to spoken commands, and summarizing large volumes of text quickly. It also includes more complex challenges like sentiment analysis, where computers determine the emotional tone behind a body of text. The Python Reddit API Wrapper (PRAW) was used to access and collect data from the Reddit platform. PRAW facilitates the extraction of Reddit posts, along with their metadata, such as post titles, creation time, and subreddit. In this study, news post titles containing the token

“controversial” were fetched from multiple subreddits: r/news, r/worldnews, and r/upliftingnews covering the period of 3 years to ensure higher relevance of the findings. Additionally, the study collects the full texts of the linked news articles using the BeautifulSoup library, which assists in extracting content from the URLs.

DATA PREPROCESSING

Prior to data analysis, preprocessing was performed to clean and prepare the data for further sentiment analysis and theme distribution. The following steps were applied during the preprocessing stage:

- a) removal of brackets and their contents from post titles using regular expressions;
- b) tokenization of post titles using the NLTK library;
- c) extraction of n-grams containing the term “controversial” using the NLTK library.

SENTIMENT ANALYSIS

The sentiment of the linked news articles was analyzed using the VADER sentiment analysis library, which is part of the NLTK (Natural Language Toolkit) library. The NLTK library is a comprehensive suite of tools for building Python programs to work with human language data, offering support for tasks such as classification, tokenization, stemming, tagging, parsing, and semantic reasoning, making it a widely used resource in the field of natural language processing. VADER is specifically designed to analyze social media text and can effectively capture sentiment in short pieces of text. The library provides polarity scores for positive, negative, and neutral sentiment, as well as a compound score representing the overall sentiment. The compound score is a metric that calculates the sum of all the lexicon ratings which have been normalized between -1 (most extreme negative) and +1 (most extreme positive). Positive sentiment: compound score ≥ 0.05 . Neutral sentiment: compound score > -0.05 and < 0.05 . Negative sentiment: compound score ≤ -0.05 . Thus, the compound score will increase as the intensity of the text increases towards positive, and, inversely, it will decrease as the intensity of the text decreases towards negative. Further on, average compound sentiment was calculated for the datasets per subreddit under investigation.

r/news and r/worldnews were selected for their broad coverage of national and international news events, making them ideal for examining a wide range of controversial issues as they unfold globally. This choice aimed to provide data on the sentiment and themes associated with controversial topics in general news discussion forums. On the other hand, r/upliftingnews was chosen as a control pool of data, as it purposefully focuses on positive news, acting as a counterbalance to the potentially negative or contentious nature of controversial topics. The hypothesis was that r/news and r/worldnews would yield different sentiment results compared to r/upliftingnews, as the latter subreddit emphasizes uplifting stories by design. This difference in focus between the subreddits was expected to be reflected in the sentiment scores, allowing for a more comprehensive understanding of the sentiment distribution in the context of controversial topics across different types of news forums.

THEME DISTRIBUTION

To identify the key themes associated with the use of “controversial” in post titles, n-grams – specific sequences of words – containing the token were examined. Specifically, the study focused on trigrams, which are strings of three uninterrupted tokens. Trigram collocations were identified using the TrigramCollocationFinder from the NLTK library. For instance, a trigram like “controversial new policy” points to a theme discussing potentially divisive new regulations, while “controversial climate change” might indicate debates around environmental issues.

STATISTICAL ANALYSIS

Descriptive statistics, such as percentages, mean, median, and standard deviation, were calculated for the sentiment scores of the linked news articles using the NumPy library – a tool for the Python programming language, designed to support complex mathematical operations and enable efficient work with large amounts of data. These statistics expose the distribution of sentiment across the collected data in addition to the overall average sentiment scores per subreddit.

CRITICAL EVALUATION OF METHODOLOGIES

The selection of methodologies for this study relied on a critical evaluation of existing approaches within the sentiment analysis and natural language processing (NLP) fields, particularly in the context of social media data analysis. Traditional sentiment analysis methods, such as VADER, offer well-proven mechanisms for capturing text sentiment, particularly suited for the brevity and informality of social media language (Hutto & Gilbert, 2014). While VADER’s lexicon-based approach is adept at handling straightforward sentiment expressions, its efficacy diminishes with complex or subtle expressions of sentiment, a limitation noted in prior research (Wankhade et al., 2022). Despite this, VADER was chosen for its proven utility in social media contexts and its ability to provide a quick, albeit broad, sentiment overview. Similarly, the use of n-grams and the Python Reddit API Wrapper (PRAW) for data collection and preprocessing addresses the need for comprehensive content analysis, allowing for the extraction of meaningful patterns and themes from large datasets. This approach is not without its weaknesses, such as the potential for oversimplification of thematic complexity and the risk of missing context-specific meanings. Acknowledging these limitations, the study combines VADER’s sentiment analysis with qualitative content analysis to ensure a balanced examination of both sentiment and thematic depth. This mixed-methods approach mitigates the methodological limitations by leveraging the strengths of each method to provide a better understanding of the sentiment and themes within controversial news posts on Reddit. By selecting these methods, the study aims to address the research questions with a methodological framework that is both robust and flexible, capable of capturing the far-from-static nature of social media discourse.

RESULTS

SENTIMENT ANALYSIS OF LINKED NEWS ARTICLES

The sentiment analysis of linked news articles containing the token “controversial” in their titles yielded the following results for the three subreddits under investigation.

A total of 321 news articles were analyzed in r/news. The distribution of sentiment scores is presented in Table 1.

TABLE 1. Distribution of sentiment scores in r/news

	Percentage	Mean	Median	SD	Average compound sentiment score
Negative Sentiment	12.55%	0.0881	0.08	0.0569	-0.1661
Positive Sentiment	7.44%	0.0744	0.063	0.0529	
Neutral Sentiment	80.01%	0.8001	0.837	0.1706	

Note: sample size n=321; positive sentiment: (compound score ≥ 0.05); neutral sentiment: (compound score > -0.05) and (compound score < 0.05); negative sentiment: (compound score ≤ -0.05)

In r/worldnews, 270 news articles were analyzed. The sentiment scores were distributed as shown in Table 2.

TABLE 2. Distribution of sentiment scores in r/worldnews

	Percentage	Mean	Median	SD	Average compound sentiment score
Negative Sentiment	11.9%	0.0745	0.073	0.0543	-0.1228
Positive Sentiment	6.72%	0.0672	0.0695	0.0561	
Neutral Sentiment	81.38%	0.8138	0.85	0.1872	

Note: sample size n=270; positive sentiment: (compound score ≥ 0.05); neutral sentiment: (compound score > -0.05) and (compound score < 0.05); negative sentiment: (compound score ≤ -0.05)

A total of 80 news articles were analyzed in r/upliftingnews. The sentiment scores distribution is presented in Table 3.

TABLE 3. Distribution of sentiment scores in r/upliftingnews

	Percentage	Mean	Median	SD	Average compound sentiment score
Negative Sentiment	9.38%	0.0737	0.0755	0.0507	0.0198
Positive Sentiment	16.35%	0.0835	0.075	0.0563	
Neutral Sentiment	74.27%	0.7427	0.816	0.2520	

Note: sample size n=80; positive sentiment: (compound score ≥ 0.05); neutral sentiment: (compound score > -0.05) and (compound score < 0.05); negative sentiment: (compound score ≤ -0.05)

The different numbers of news articles analyzed in each subreddit result from our method of data collection, which was based on a specific timeframe of 3 years, rather than aiming for a predetermined number of posts or articles. Given this timeframe, each subreddit naturally had a different number of posts titled with “controversial” due to the varying posting activity within their communities. The study’s analytical approach accounts for these differences by employing relative measures (such as percentages) in the sentiment analysis and theme distribution exploration.

THEME DISTRIBUTION

Based on the retrieved trigrams from the subreddits, several themes were identified. The distribution of these themes is as follows.

A significant proportion of the trigrams were related to Politics and Legislation, accounting for 26.32% of the total. The Technology and Surveillance theme constituted 12.28% of the trigrams. Trigrams related to Social Issues and Controversies made up 10.53% of the total. The Health and Medicine theme accounted for 8.77% of the trigrams. Trigrams in the Environment and Energy theme constituted 7.02% of the total. The remaining 35.08% of the trigrams were distributed among various other themes such as education, religion, business, entertainment, military, and international relations. Table 4 summarizes data results retrieved for theme distribution with examples.

TABLE 4. Distribution of themes associated with the “controversial” token across three subreddits

Theme	Percentage in the sample	Examples of retrieved trigrams
Politics and Legislation	26.32%	controversial election-law case controversial legislation against controversial Georgia voting controversial copyright rules controversial tax reform controversial judicial overhaul controversial Brexit protocol controversial immigration policy controversial gun-control measures controversial finance regulations
Technology and Surveillance	12.28%	controversial facial recognition controversial facial-recognition technology controversial Clearview AI controversial mass surveillance controversial spyware technology controversial data privacy regulation controversial internet censorship controversial AI-driven policing controversial biometric tracking controversial drone surveillance
Social Issues and Controversies	10.53%	controversial virginity tests controversial cuties movie controversial arrest of controversial asylum detention controversial BBC documentary controversial gender-neutral bathrooms controversial racial profiling controversial pay-gap debate controversial police brutality controversial maternity bill
Health and Medicine	8.77%	controversial new drug controversial opioid treatment controversial third-dose vaccination controversial vaccine bills controversial Covid rules controversial stemcell research controversial assisted suicide

		controversial mental health
		controversial abortion restrictions
		controversial GMO food
Environment and Energy	7.02%	controversial Farallon islands
		controversial pebble mine
		controversial wolf cull
		controversial mining project
		controversial drilling operations
		controversial pipeline construction
		controversial NP plant
		controversial deforestation policy
		controversial fracking regulations
		controversial climate-change policy
Other	35.08%	controversial Mohammed cartoons
		controversial scene depicting
		controversial pastor Tony
		controversial military program
		controversial university admissions
		controversial religious conversion
		controversial corporate tax
		controversial casting choice
		controversial military intervention
		controversial diplomatic relations
		controversial curriculum changes
		controversial celebrity endorsement
		controversial trade deal
		controversial standardized testing
		controversial CEO compensation

Transitioning from the quantitative analysis, the following Discussion will interpret these findings, specifically addressing the implications of the sentiment and theme distribution results.

DISCUSSION

INTERPRETATION OF SENTIMENT ANALYSIS RESULTS

Study results have revealed the sentiment associated with controversial topics in news-related Reddit posts. By selecting r/news and r/worldnews, the study was able to examine the general sentiment surrounding controversial subjects in popular news forums. In contrast, turning to r/upliftingnews as a control pool of data proved useful, as study results have confirmed our expectation that it would yield different sentiment results compared to the other two subreddits.

The sentiment analysis results indicate that the news posts in r/news and r/worldnews are predominantly neutral, with 80.01% and 81.38% of the content having neutral sentiment, respectively. However, the average compound sentiment scores were negative for both r/news (-0.1661 \leq -0.05) and r/worldnews (-0.1228 \leq -0.05), suggesting that there is still a negative undertone in the framing of controversial topics in these subreddits. In contrast, r/upliftingnews, which focuses on positive news stories, had a noticeably different sentiment profile. Although most of the content (74.27%) still had a neutral sentiment, the average compound sentiment score was positive (0.0198 \geq 0.05). This finding indicates that controversial topics in r/upliftingnews are framed in a more positive light, thus pointing to the variation in the framing of controversial topics between different subreddits. The negative sentiments observed in r/upliftingnews, despite

its overall positive orientation, could be attributed to two key factors. First, the term “controversial” inherently suggests a divergence of views, which can evoke a broad spectrum of emotional responses, including negative ones, even when the overarching theme is uplifting. For example, a post discussing a controversial but ultimately successful environmental conservation effort might carry a negative sentiment due to the contentious nature of the debate, despite the positive outcome. Furthermore, the framing of a story as “controversial” within r/upliftingnews might reflect the challenges or obstacles overcome to achieve a positive result, thereby introducing elements of conflict or adversity into the narrative. These stories, while ultimately affirming, may initially present negative aspects to accentuate the contrast with the positive resolution.

While the compound sentiment score provides a snapshot of overall sentiment within each subreddit, a closer examination through mean, median, and standard deviation (SD) reveals the depth of sentiment distribution and variability. Specifically, the mean sentiment scores indicate the average sentiment intensity, with r/news and r/worldnews showing higher mean negative sentiment scores (0.0881 and 0.0745, respectively) than r/upliftingnews (0.0737), suggesting a stronger presence of negative sentiment in these forums. In simpler terms, this indicates that on average, discussions in r/news and r/worldnews are more negatively toned compared to those in r/upliftingnews. The median values, representing the midpoint of sentiment scores, further clarify sentiment concentration. For negative sentiment, the median in r/news is 0.08, higher than r/worldnews’s 0.073 and r/upliftingnews’s 0.0755, meaning that, on average, discussions in r/news tend to be closer to a negative sentiment than those in r/worldnews and r/upliftingnews. The SD quantifies sentiment score dispersion around the mean; higher SDs in r/news (0.0569) and r/worldnews (0.0543) compared to r/upliftingnews (0.0507) indicate a wider variability of negative sentiment within these subreddits. In other words, the SD, or standard deviation, tells us how much the sentiment varies from the average, and in r/news and r/worldnews, sentiments are more spread out, showing a broader range of negative feelings compared to r/upliftingnews. Collectively, these statistics affirm that r/news and r/worldnews not only harbor more pronounced negative sentiment but also exhibit greater sentiment dispersion than r/upliftingnews.

Importantly, the predominance of neutral sentiment in the linked text articles suggests that most of the articles are fact-based and present information without strong emotional connotations. This observation can be instrumental in dissecting how controversial topics are framed in news articles on Reddit, indicating that authors tend to present information in a more neutral manner. However, as is shown in study results, the overall compound score, which captures positive, negative, and neutral sentiment, can still be negative even if 70 to 80% of the dataset is neutral. This is because the compound score is a measure of the overall sentiment across all articles in a subreddit. In our case, although the majority of the articles may have neutral sentiment, the remaining negative sentiment appeared strong enough to ultimately influence the overall compound score to a statistically significant extent. Thus, in the case of r/news and r/worldnews, the negative compound scores suggest that the negative sentiment in the remaining articles is strong enough to offset the neutral sentiment, resulting in a negative overall compound score. This finding demonstrates that there is a negative undertone in the framing of controversial topics in these subreddits, despite the high percentage of neutral sentiment articles.

This finding is particularly significant as it contrasts with existing literature that primarily categorizes online sentiment in binary terms of positive versus negative (Taj et al., 2019; Shapiro et al., 2022; Yan et al., 2021). While these and other studies concerned with sentiment analysis emphasize the prevalence of explicit sentiment expressions, our research has identified the subtlety of sentiment expression in online news discourse, where neutral phrasing can coexist with a

palpable negative sentiment. This phenomenon puts into the focus the influence of negative framing in ostensibly neutral discussions, suggesting that even a limited presence of negative sentiment can significantly color the overall discourse, a subtlety not fully captured in traditional sentiment analysis approaches.

As such, this finding can have considerable implications in the context of both existing and future research and practice. For example, comparing our results to those of Shapiro et al. (2022) and Taj et al. (2019), who primarily focus on the aggregate impact of positive and negative sentiments, our study specifically points to the critical role of neutral sentiment in shaping discourse framing, whereby neutral sentiment does not equate to a neutral impact on discourse. Furthermore, our analysis diverges from Yan et al. (2021), who explore sentiment in the context of health crises framing in the news, primarily focusing on explicit sentiment expressions. In contrast, our study reveals that the framing of controversial topics, even when predominantly neutral, may carry an underlying negative bias. This suggests implications for content creation practices, audience engagement, and sentiment analysis methodology, as will be discussed further.

INTERPRETATION OF THEME DISTRIBUTION RESULTS

The theme distribution analysis provides a deeper understanding of the types of controversial topics discussed across the three subreddits. The most predominant theme associated with the term “controversial” in the post titles across the three subreddits was Politics and Legislation (26.32%), followed by Technology and Surveillance (12.28%), Social Issues and Controversies (10.53%), Health and Medicine (8.77%), and Environment and Energy (7.02%). The remaining themes collectively accounted for 35.08% of the trigrams. This distribution highlights the significant role that politics and legislation play in shaping controversial discussions on social media platforms.

The theme distribution analysis not only provides a deeper appreciation of the types of controversial topics discussed across the three subreddits, but also showcases the various aspects of society that are prone to controversy and debate. As mentioned earlier, Politics and Legislation emerged as the most predominant theme (26.32%), reflecting the highly polarized and contested nature of political discourse in the contemporary social media landscape. The prominence of this theme suggests that users are actively engaging with and discussing policy-related matters, which can significantly impact their lives and society at large. This finding resonates with a growing body of research that points out the role of social media as a forum for political discourse and civic engagement. Studies such as Boulianne’s (2015) have demonstrated a positive correlation between social media use and political participation, indicating that platforms like Reddit can serve as important spaces for political mobilization and discussion. This is further supported by Valenzuela et al.’s (2019) work, which found that engagement with social media leads to increased political engagement among users, particularly in terms of information seeking and opinion expression. Moreover, the research by Gil de Zúñiga et al. (2012) highlights the concept of “social media as a public sphere”, where users not only consume content but actively participate in political discussions, thereby contributing to the public discourse on policy-related matters. Additionally, recent studies focusing on the impact of digital platforms on policy discussions reveal that social media significantly shapes the framing and perception of political and legislative issues (Margetts et al., 2015). These scholarly contributions collectively affirm the significance of our observation, exposing the interconnection between social media discourse and the broader socio-political context that drives the prominence of policy-related matters in online communities.

The second most predominant theme, Technology and Surveillance (12.28%), highlights the growing concerns and debates surrounding privacy, data protection, and the ethical

implications of technological advancements. Building on past research that points out the pervasive impact of technology on society, such as the works of Fuchs (2017) and Zuboff (2019), our findings reflect similar concerns within the Reddit community. The prominence of Technology and Surveillance themes in our study, especially concerning privacy and data ethics, echoes the global conversation on digital rights and tech governance. For example, the debates on the balance between security and privacy in the use of surveillance technologies, as explored in the European Union's General Data Protection Regulation (GDPR), are mirrored in the texts analyzed, indicating a widespread public engagement with these critical issues.

The Social Issues and Controversies theme (10.53%) underscores the wide range of societal concerns that spark discussions and debates on social media platforms. These issues may include race, gender, immigration, and other topics that are often deeply rooted in cultural and historical contexts. The prevalence of this theme suggests that social media platforms serve as crucial spaces for raising awareness and fostering dialogue around these critical issues. These topics' presence on social media demonstrates platforms like Reddit's role in amplifying discussions on societal issues, akin to Castells' (2015) observation of social networks as spaces for political and social movements.

The Health and Medicine theme (8.77%) demonstrates the significance of healthcare-related controversies in public discourse, especially considering the recent global health crisis and debates around vaccination, healthcare access, and the role of the pharmaceutical industry (Benrouba & Boudour, 2023; Fernández-Luque & Bau, 2015; Yan et al., 2021). This theme's prominence also points to the importance of accurate and reliable information dissemination regarding health and medicine to ensure informed decision-making by the public.

The Environment and Energy theme (7.02%) mirrors the growing public concern over environmental sustainability, climate change, and the pursuit of renewable energy sources. This thematic prominence bears testimony to an escalating dialogue within online communities about the pressing need to combat environmental degradation and transition towards cleaner energy practices (Taufek et al., 2021). For instance, discussions often revolve around topics such as the Paris Agreement's goals for reducing carbon emissions, the alarming rates of deforestation in the Amazon rainforest, or the potential of solar and wind energy to replace fossil fuels. The increase in posts related to innovative recycling technologies, urban sustainability projects like green roofing, and debates over the efficacy of electric vehicles exemplifies the diverse range of topics within this theme.

The theme distribution results highlight the interconnectedness of various themes and their potential to influence one another. For instance, Politics and Legislation may have direct implications on Technology and Surveillance, Health and Medicine, and Environment and Energy, as policies and regulations often govern these sectors. This interconnectedness emphasizes the importance of fostering interdisciplinary dialogues and collaborations in addressing complex, controversial issues.

Another valuable observation concerns the potential impact of current events and evolving societal priorities on the theme distribution. For example, during times of heightened political activity, such as election seasons, the prominence of Politics and Legislation may increase. Similarly, global health crises or environmental disasters may elevate the significance of Health and Medicine or Environment and Energy themes. Analyzing the theme distribution over time could reveal the ways in which public discourse evolves in response to changing global dynamics.

Lastly, the theme distribution analysis can also give exposure to the different types of controversial topics that may engage diverse demographics and user groups on social media

platforms. Understanding the preferences and concerns of various user segments can help tailor communication strategies and create targeted interventions that address the specific needs and interests of these groups, fostering more inclusive and effective discussions around controversial topics.

Thus, the theme distribution analysis offers actionable insights into the various aspects of society that generate controversy and debate on social media platforms. These findings can inform future research on social media discourse and guide discussions and policies related to controversial topics in online communities, ensuring a more inclusive and constructive dialogue.

IMPLICATIONS OF STUDY FINDINGS

The key finding of our research – that predominantly neutral sentiment in the discussion of controversial topics on platforms like Reddit can mask an underlying negative undertone – has far-reaching implications for both academic research and practical applications in content creation and media strategy.

Content creation practices. Content creators, particularly those involved in news dissemination and discussion facilitation on social media platforms, need to be acutely aware of the subtle influences of sentiment framing. While the aim may be to present information neutrally, the inadvertent skew towards negative sentiment, as discovered by our findings, can profoundly affect content tone. This calls for a thoughtful content framing, where the potential for negative bias, even in neutral presentations, is carefully managed to avoid unintended perceptions. Additionally, the identified themes can be useful for practitioners interested in understanding the types of controversial issues that generate discussions on social media platforms. This knowledge can also be beneficial for researchers aiming to study public opinion, online behavior, or the impact of controversial topics on user engagement.

Enhanced sentiment analysis tools. For practitioners in the field of natural language processing (NLP) and sentiment analysis, our study points to the need for advanced tools that can detect and quantify the impact of subtle negative biases within predominantly neutral content. This could lead to the development of more sophisticated and – importantly – more widely used and recognized – sentiment analysis algorithms that are better equipped to understand and analyze complex human language, in which neutrality can be overwhelmingly offset by even marginal negative undertones.

Content moderation strategies. Social media platforms and news organizations could utilize findings from our study to refine their content moderation strategies. By recognizing that neutral content may still carry a negative undertone, moderators can better identify and address potentially divisive or misleading framing of controversial topics, thereby supporting a more balanced and constructive online discourse.

Audience engagement and trust. Understanding the impact of sentiment on discourse allows content creators to engage more effectively with their audience. By avoiding unintentional negative biases in content that is intended to be neutral, creators can build greater trust and credibility with their audience. This is particularly important in contexts where neutrality and objectivity are valued, such as in news reporting and informational content related to sensitive or controversial topics.

Thus, our findings challenge the conventional dichotomy of positive versus negative sentiment analysis, introducing a critical perspective on the role of neutral sentiment in shaping online discourse. This perspective may offer practical guidance for content creators, strategists and scholars aiming to navigate the complexities of sentiment in digital communication effectively.

CONCLUSION

The study aimed to fill a significant gap in the literature regarding the exploration of platform-specific news sentiment dynamics with typological content differentiation. To address this gap, we examined the sentiment of news articles on Reddit that contain the token “controversial” in their titles and identified the key themes associated with its use. The research questions focused on identifying and interpreting the sentiment distribution and the prevalence of themes in the context of controversial topics. To address these questions, a mixed-methods NLP approach was employed, combining Python-instrumented sentiment analysis using the VADER sentiment analysis library and a qualitative content analysis using trigrams to identify and categorize themes.

The analysis yielded two main sets of findings. First, the sentiment analysis revealed that most news posts in r/news and r/worldnews exhibit a neutral sentiment, though with a negative undertone as indicated by the average compound sentiment scores. This contrasts with r/upliftingnews, where a more positive sentiment was observed, albeit still with a significant portion of neutral sentiment. This suggests that while the presentation of information may appear neutral, there exists a subtle negative framing of controversial topics on Reddit. Critically, this exposes the importance of considering the interconnection between neutral and negative sentiments in understanding online news discourse. It suggests that future research should incorporate sentiment analysis techniques capable of capturing the complexity of sentiment interaction, particularly in contexts where neutral presentation coexists with underlying negative biases.

Second, the theme distribution analysis revealed that Politics and Legislation was the most predominant theme across the three subreddits, followed by Technology and Surveillance, Social Issues and Controversies, Health and Medicine, and Environment and Energy. This distribution highlights the various aspects of society that generate controversy and debate on social media platforms, registering the nature of public discourse on these topics.

Having suggested valuable data and observations on sentiment and themes around controversial topics on Reddit, this study, however, is not without limitations. The selected subreddits represent a specific segment of Reddit and may not encompass all relevant content. Additionally, the VADER tool used for sentiment analysis, while effective, may not capture the full subtlety of language, such as sarcasm or mixed emotions. Future research could broaden the scope to include more subreddits and employ diverse analytical tools to better appreciate sentiment in online discourse.

The key implications of our research are manifold. For academics, it signifies the importance of elaborate sentiment analysis beyond binary classifications of positive and negative, suggesting that neutral sentiment can still carry significant underlying tones. For practitioners, especially those involved in content creation and moderation on social media platforms, our findings emphasize the need for careful consideration of sentiment framing to avoid unintended biases in the presentation of controversial topics. Moreover, the identified themes have elicited the types of issues that engage users and provoke discussion, which can inform content strategy and moderation policies aimed at incentivizing constructive discourse.

REFERENCES

- Agapova, E. A., & Grischechko, E. G. (2016). Censorship as a factor of information warfare. *Russian Linguistic Bulletin*, 3(7), 43-44. <https://doi.org/10.18454/RULB.7.06>
- Agur, C., & Frisch, N. (2019). Digital disobedience and the limits of persuasion: Social media activism in Hong Kong's 2014 Umbrella Movement. *Social Media and Society*, 5(1). <https://doi.org/10.1177/2056305119827002>
- Aldous, K. K., An, J., & Jansen, B. J. (2023). What really matters? Characterising and predicting user engagement of news postings using multiple platforms, sentiments and topics. *Behaviour & Information Technology*, 42(5), 545-568. <https://doi.org/10.1080/0144929X.2022.2030798>
- Alkhamash, R. (2021). The social media framing of gender pay gap debate in American women's sport: A linguistic analysis of emotive language. *Training, Language and Culture*, 5(4), 22-35. <https://doi.org/10.22363/2521-442X-2021-5-4-22-35>
- Ash, E., Xu, Y., Jenkins, A., & Kumanyika, C. (2019). Framing use of force: An analysis of news organizations' social media posts about police shootings. *Electronic News*, 13(2), 93-107. <https://doi.org/10.1177/1931243119850239>
- Azhari, A., & Fang, X. (2018). Social media applications framework: A lexical analysis of users online reviews. *International Journal of Services and Standards*, 12(2), 140-162. <https://doi.org/10.1504/IJSS.2018.091850>
- Aziz, J. B., & Hashim, F. (2021). Rhetoric of food authenticity and national identity in the new media. *GEMA Online Journal of Language Studies*, 21(2), 253-272. <https://dx.doi.org/10.17576/gema-2021-2102-14>
- Bednarek, M. (2006). *Evaluation in media discourse: Analysis of a newspaper corpus*. A&C Black.
- Belcastro, L., Branda, F., Cantini, R., Marozzo, F., Talia, D., & Trunfio, P. (2022). Analyzing voter behavior on social media during the 2020 US presidential election campaign. *Social Network Analysis and Mining*, 12(1). <https://doi.org/10.1007/s13278-022-00913-9>
- Benrouba, F., & Boudour, R. (2023). Emotional sentiment analysis of social media content for mental health safety. *Social Network Analysis and Mining*, 13(1). <https://doi.org/10.1007/s13278-022-01000-9>
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192-205. <https://doi.org/10.2139/ssrn.1528077>
- Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, Communication & Society*, 18(5), 524-538. <https://doi.org/10.1080/1369118X.2015.1008542>
- Castells, M. (2015). *Networks of outrage and hope: Social movements in the Internet age*. John Wiley & Sons.
- Chau, D., & Lee, C. (2021). "See you soon! ADD OIL AR!": Code-switching for face-work in edu-social Facebook groups. *Journal of Pragmatics*, 184, 18-28. <https://doi.org/10.1016/j.pragma.2021.07.019>
- Fernández-Luque, L., & Bau, T. (2015). Health and social media: Perfect storm of information. *Healthcare Informatics Research*, 21(2), 67-73. <https://doi.org/10.4258/hir.2015.21.2.67>
- Fuchs, T. (2017). *Ecology of the brain: The phenomenology and biology of the embodied mind*. Oxford University Press.

- Gil de Zúñiga, H., Jung, N., & Valenzuela, S. (2012). Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, 17(3), 319-336. <https://doi.org/10.1111/j.1083-6101.2012.01574.x>
- Grašič, T. (2022). The writing style in travel blogs. *Folia Linguistica et Litteraria*, 13(41), 187-210. <https://doi.org/10.31902/fl.41.2022.9>
- Grishechko, E. G. (2023a). Emojis as nonverbal cues in online communication: Perspectives on conflict resolution and misunderstanding prevention. In *Proceedings of 10th SWS International Scientific Conference on Arts And Humanities – ISCAH 2023* (Vol. 23, No. 1). SGEM World Science (SWS) Scholarly Society. <https://doi.org/10.35603/sws.iscah.2023/s11.12>
- Grishechko, E. G. (2023b). Language and cognition behind simile construction: A Python-powered corpus research. *Training, Language and Culture*, 7(2), 80-92. <https://doi.org/10.22363/2521-442X-2023-7-2-80-92>
- Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Huynh, D., Audet, G., Alabi, N., & Tian, Y. (2021, December 15018). Stock price prediction leveraging Reddit: The role of trust filter and sliding window. In *Proceedings of 2021 IEEE International Conference on Big Data* (pp. 1054-1060). IEEE. <https://doi.org/10.1109/BigData52589.2021.9671412>
- Johannessen, M. R. (2015, August 30 – September 2). Please like and share! A frame analysis of opinion articles in online news. In *Proceedings of the 7th IFIP 8.5 International Conference* (pp. 15-26). Springer. https://doi.org/10.1007/978-3-319-22500-5_2
- Kasperuniene, J., & Zydziunaite, V. (2019). A systematic literature review on professional identity construction in social media. *Sage Open*, 9(1). <https://doi.org/10.1177/2158244019828847>
- Kavitha, M., Naib, B. B., Mallikarjuna, B., Kavitha, R., & Srinivasan, R. (2022, April 28-29). Sentiment analysis using NLP and machine learning techniques on social media data. In *Proceedings of the 2nd International Conference on Advance Computing and Innovative Technologies in Engineering* (pp. 112-115). IEEE. <https://doi.org/10.1109/ICACITE53722.2022.9823708>
- Kralj Novak, P., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PloS One*, 10(12), e0144296. <https://doi.org/10.1371/journal.pone.0144296>
- Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), 188-206. <https://doi.org/10.1080/23808985.2021.1976070>
- Kumar, N., Nagalla, R., Marwah, T., & Singh, M. (2018). Sentiment dynamics in social media news channels. *Online Social Networks and Media*, 8, 42-54. <https://doi.org/10.1016/j.osnem.2018.10.004>
- Lai, C. H., & Fu, J. S. (2021). Exploring the linkage between offline collaboration networks and online representational network diversity on social media. *Communication Monographs*, 88(1), 88-110. <https://doi.org/10.1080/03637751.2020.1869797>
- López-Rabadán, P. (2021). Framing studies evolution in the social media era. Digital advancement and reorientation of the research agenda. *Social Sciences*, 11(1). <https://doi.org/10.3390/socsci11010009>

- Machavarapu, A. (2022). Reddit sentiments effects on stock market prices. In V. Bhateja, S. C. Satapathy, C. M. Travieso-Gonzalez, & T. Adilakshmi, T. (Eds.), *Smart intelligent computing and applications* (pp. 75-84). Springer. https://doi.org/10.1007/978-981-16-9669-5_7
- Malyuga, E. N., & Akopova, A. S. (2021). Precedence-setting tokens: Issues of classification and functional attribution. *Training, Language and Culture*, 5(4), 65-76. <https://doi.org/10.22363/2521-442X-2021-5-4-65-76>
- Margetts, H. Z., John, P., Hale, S. A., & Reissfelder, S. (2015). Leadership without leaders? Starters and followers in online collective action. *Political Studies*, 63(2), 278-299. <https://doi.org/10.1111/1467-9248.12075>
- Miyake, K. (2007). How young Japanese express their emotions visually in mobile phone messages: A sociolinguistic analysis. *Japanese Studies*, 27(1), 53-72. <https://doi.org/10.1080/10371390701268646>
- Mostafa, M. M. (2019). Clustering halal food consumers: A Twitter sentiment analysis. *International Journal of Market Research*, 61(3), 320-337. <https://doi.org/10.1177/1470785318771451>
- Oh, H., Goh, K. Y., & Phan, T. Q. (2023). Are you what you tweet? The impact of sentiment on digital news consumption and social media sharing. *Information Systems Research*, 34(1), 111-136. <https://doi.org/10.1287/isre.2022.1112>
- Omar, N., & Al-Tashi, Q. (2018). Arabic nested noun compound extraction based on linguistic features and statistical measures. *GEMA Online Journal of Language Studies*, 18(2), 93-107. <https://doi.org/10.17576/gema-2018-1802-07>
- Păvăloaia, V. D., Teodor, E. M., Fotache, D., & Daniieș, M. (2019). Opinion mining on social media data: Sentiment analysis of user preferences. *Sustainability*, 11(16). <https://doi.org/10.3390/su11164459>
- Schoenebeck, S., Lampe, C., & Triêu, P. (2023). Online harassment: Assessing harms and remedies. *Social Media and Society*, 9(1). <https://doi.org/10.1177/20563051231157297>
- Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, 228(2), 221-243. <https://doi.org/10.24148/erwp2017-01>
- Shifman, L. (2013). *Memes in digital culture*. MIT Press.
- Sibul, V. V., Vetrinskaya, V. V., & Grishechko, E. G. (2019). Study of precedent text pragmatic function in modern economic discourse. In E. N. Malyuga (Ed.), *Functional approach to professional discourse exploration in linguistics* (pp. 131-163). Springer. https://doi.org/10.1007/978-981-32-9103-4_5
- Splendiani, S., & Capriello, A. (2022). Crisis communication, social media and natural disasters: The use of Twitter by local governments during the 2016 Italian earthquake. *Corporate Communications*, 27(3), 509-526. <https://doi.org/10.1108/CCIJ-03-2021-0036>
- Taj, S., Shaikh, B. B., & Meghji, A. F. (2019, January 30-31). Sentiment analysis of news articles: A lexicon-based approach. In *Proceedings of the 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-5). IEEE. <https://doi.org/10.1109/ICOMET.2019.8673428>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153. <https://doi.org/10.1080/21670811.2017.1360143>

- Taufek, T. E., Nor, N. F. M., Jaludin, A., Tiun, S., & Choy, L. K. (2021). Public perceptions on climate change: A sentiment analysis approach. *GEMA Online Journal of Language Studies*, 21(4), 209-233. <https://doi.org/10.17576/gema-2021-2104-11>
- Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism*, 7(6), 802-823. <https://doi.org/10.1080/21670811.2019.1623701>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>
- Waddell, T. F. (2020). The authentic (and angry) audience: How comment authenticity and sentiment impact news evaluation. *Digital Journalism*, 8(2), 249-266. <https://doi.org/10.1080/21670811.2018.1490656>
- Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780. <https://doi.org/10.1007/s10462-022-10144-1>
- Wu, G., & Pan, C. (2022). Audience engagement with news on Chinese social media: A discourse analysis of the People's Daily official account on WeChat. *Discourse & Communication*, 16(1), 129-145. <https://doi.org/10.1177/175048132111026567>
- Wu, H., Bakar, K. A., Jaludin, A., & Awal, N. M. (2022). Sentiment analysis of China-related news in *The Star Online* newspaper. *GEMA Online Journal of Language Studies*, 22(3), 155-175. <https://doi.org/10.17576/gema-2022-2203-09>
- Yan, C., Law, M., Nguyen, S., Cheung, J., & Kong, J. (2021). Comparing public sentiment toward COVID-19 vaccines across Canadian cities: Analysis of comments on Reddit. *Journal of Medical Internet Research*, 23(9), e32685. <https://doi.org/10.2196/32685>
- Zappavigna, M. (2018). *Searchable talk: Hashtags and social media metadiscourse*. Bloomsbury Publishing.
- Žitnik, S., Blagus, N., & Bajec, M. (2022). Target-level sentiment analysis for news articles. *Knowledge-Based Systems*, 249, 108939. <https://doi.org/10.1016/j.knosys.2022.108939>
- Zubbir, N., Dass, L. C., & Ahmad, N. (2021). Analysis on adjective Suki and its co-occurrences in Japanese YouTube's comment. *Pertanika Journal of Social Sciences & Humanities*, 29(2), 1357-1374. <https://doi.org/10.47836/pjssh.29.2.32>
- Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum*, 28(1), 10-29. <https://doi.org/10.1177/1095796018819461>

ABOUT THE AUTHOR

Elizaveta G. Grishechko, Cand. Sci. (Linguistics), Senior Lecturer in the Foreign Languages Department, Faculty of Economics, RUDN University, Russia. Research interests cover the issues of cultural linguistics, corpus linguistics, sentiment analysis, natural language processing and the language of academic writing.