Energy Policy in Flux: An NLP Approach to Media Discourse Before and After the Russia-Ukraine War

Priscila Rebeca Stisman

April 29th, 2025

Abstract

This paper explores shifts in the discourse surrounding energy policy by applying natural language processing (NLP) techniques to *New York Times* articles published between 2021 and 2023. Using dictionary-based analysis, topic modeling methods—including Latent Dirichlet Allocation (LDA) and Structural Topic Modeling (STM)—as well as a time series regression analysis, I examine whether the prominence and framing of energy-related issues changed after the onset of the Russia-Ukraine war. The results indicate that energy-related topics became significantly more prominent following the conflict, though their salience gradually declined over time, suggesting a normalization of coverage or a shift in media focus.

Keywords: *Energy Policy, NLP, Topic Modeling, Sentiment Analysis*

1. Introduction

In recent years, energy policy has become central to public discourse. Topics like renewable energy, fossil fuels, and natural gas—once technical or niche—are now widely debated, driven by growing concern over climate change and crises such as the Russia-Ukraine war, which exposed vulnerabilities in global energy systems.

This heightened attention is not without controversy. While many support urgent investment in clean energy and security, others question the urgency of climate action or prioritize economic challenges like poverty and inequality. These diverging perspectives make it important to examine how public discourse around energy has evolved in recent years.

One valuable way to trace such shifts in public perception is through media reporting [3]. Media outlets play a crucial role in shaping the policy agenda by connecting policymakers and the public. In this context, analyzing media coverage offers a useful proxy for understanding societal attention to and concern about energy-related issues.

The Russia-Ukraine war, which began on February 24, 2022, marked a potential turning point in global energy policy. Given Europe's significant dependence on Russian oil and gas, the conflict triggered a re-evaluation of energy strategies worldwide. In its aftermath, several scholars have observed increased support for clean energy initiatives across much of the political spectrum [4]. Beyond clean energy, the war has had broader implications for energy materiality—from climate impacts to fertilizer supply—prompting a reassessment of how energy systems are studied and understood [1].

In this paper, I aim to examine whether there has been a significant shift in the media discourse surrounding energy since the onset of the Russia-Ukraine war. This conflict has had profound implications for global energy markets and security, potentially reshaping both the way energy issues are discussed in the media and the level of attention society pays to them. For this analysis, I will focus on The New York Times, one of the most influential newspapers in the world. With its extensive international readership and established role in shaping the global news agenda, The New York Times provides a valuable lens through which to assess changing narratives and shifting public attention on energy-related matters.

2. Data

For this research, I analyze New York Times articles from 2021 to 2023, filtered by energy-related keywords. The dataset, obtained through the NYT API, includes over two million articles published between 2000 and 2025, each containing detailed metadata that is available in real-time upon publication.

Focusing on the period surrounding the onset of the Russia-Ukraine war, I retain 126,109 articles from 2021 to 2023. From each article, I extract four key features: the abstract, headline, lead paragraph, and publication date.

To examine shifts in energy policy discourse during this period, I apply additional filters using energy-related keywords, resulting in a final subset of 18,293 articles.

3. Methods

The analysis begins with text preprocessing, transforming raw text into a suitable corpus by removing punctuation, numbers, stopwords, and applying stemming. A document-feature matrix is then created, removing rare and overly common terms to focus on meaningful vocabulary.

Next, sentiment analysis is performed on energy policy articles to assess shifts in tone before and after the war. Latent Dirichlet Allocation (LDA) is applied to identify key themes within energy-related content, followed by Interrupted Time Series (ITS) regression to evaluate the war's impact on the salience of energy topics in media coverage. Finally, Structural Topic Modeling (STM) is employed to track the evolution of energy discourse over time.

4. Results

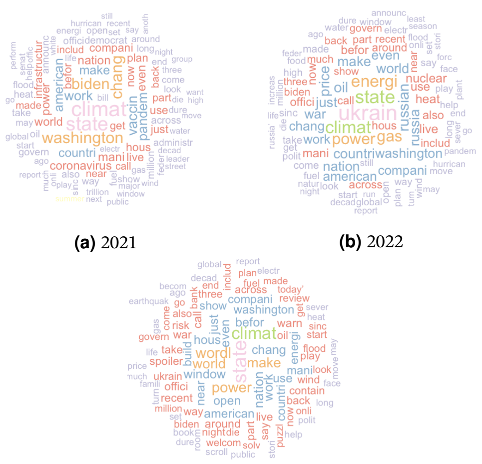
Using the pre-processed text, I generate word clouds for each year to visually capture the most prominent terms in the energy discourse (see Figure 1). In addition, I extract the top 10 most frequent words from each year to identify salient themes and possible changes in the language used around energy topics over time.

In 2021, the most frequent terms reflect focus on U.S. politics and COVID-19 recovery. In 2022, a shift occurs, with terms like "Ukraine," "energy," "climate," "gas," "oil," "price," "Russian," and "war" rising, reflecting heightened concerns over energy security and geopolitical tensions due to the Russia-Ukraine war. By 2023, only two energy-related terms—"climate" and "power"—remain prominent, suggesting a decrease in media focus on energy as the crisis's initial impact waned.

Year	Top 10 Words
2021	climat, state, chang, washington, biden, vaccin, work, pandem, american, make
2022	ukrain, state, climat, energi, power, gas, russian, oil, price, war
2023	state, climat, wordl, power, make, world, nation, even, work, window

4.1. Dictionary-Based Analysis

I conduct sentiment analysis on article headlines using the VADER (Valence Aware Dictionary and sEntiment Reasoner) algorithm, which assigns a compound sentiment score ranging from -1 (most



(a) 2021 **(b)** 2022

(c) 2023

Figure 1. Word Clouds

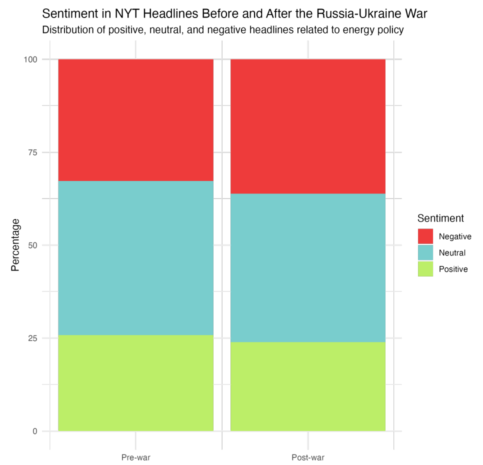


Figure 2. Sentiment Analysis of Energy Policy Discourse Pre and Post War

negative) to +1 (most positive). Based on these scores, each headline is classified as positive, neutral, or negative.

To focus more specifically on coverage related to policy, I further refine the energy corpus by filtering it using policy-related keywords. This results in a final energy-policy corpus of 6,664 articles. I then compare the distribution of sentiment before and after the onset of the Russia-Ukraine war.

As shown in Figure 2, negative sentiment increases after the war, with negative headlines rising from 32% pre-war to 36% post-war. This shift reflects the global energy crisis triggered by the conflict, marked by rising oil and gas prices and supply concerns. However, increased negative sentiment does not indicate opposition to energy policy but rather heightened anxiety and urgency, as seen in headlines like *"Oil prices climb as Ukraine crisis deepens*" which reflect concern rather than a policy stance.

4.2. Latent Dirichlet Allocation

To examine how energy-related news coverage shifted following the Russia-Ukraine war, I perform topic modeling on the whole energy corpus using the Latent Dirichlet Allocation (LDA) algorithm [2]. This analysis aims to identify the dominant topics in energy discourse before and after the war. To select the optimal number of topics, I evaluate coherence and perplexity scores[1], which indicate K* = 5. The results reveals a clear shift in media coverage from pre-war to post-war. Prior to the war, the discourse was relatively diverse, encompassing topics such as climate change, domestic U.S. politics, and recovery from the COVID-19 pandemic. Following the onset of the war, however, the focus pivoted sharply toward energy sources, geopolitics, and the conflict itself (see Figure 3).

Notably, topics related to nuclear energy, gas, and oil gained prominence, reflecting concerns over energy supply and geopolitical ramifications. Keywords like "nuclear," "gas," "oil," "power," and "war" became central, marking a shift from earlier discussions on domestic issues and climate policy to urgent questions about international energy systems and security.

Overall, the findings suggest that the war reoriented media narratives, placing gas, oil, and nuclear power at the forefront of global energy policy debates.[2]

4.3. Interrupted Time Series Regression

To better understand how the Russia–Ukraine conflict shaped public discourse on energy policy, I construct an energy salience score that quantifies the prominence of energy-related language in news articles. Specifically, this score is calculated as the proportion of energy-related words, e.g. "oil", "gas", "fuel", "fossil", "nuclear", that appear in the headline, abstract, and lead paragraph of each article, relative to the total number of words in those sections. The goal is to capture how central energy topics are in the framing of the news over time.

I begin the analysis by estimating a baseline regression that evaluates whether the war had a statistically significant effect on the energy salience of news articles. This initial model includes a simple binary indicator for whether an article was published after the start of the war. By comparing the average energy scores before and after the war, this approach allows me to detect whether there was a discrete shift in the importance of energy topics following the conflict.

$$EnergyScore = \beta_0 + \beta_1 War \qquad (1)$$

Variable	Energy Score
Intercept	3.2066 ***
	(0.039)
War	-0.05329
	(0.0535)
F(1,9295)= 0.9937 p-value: 0.3189	

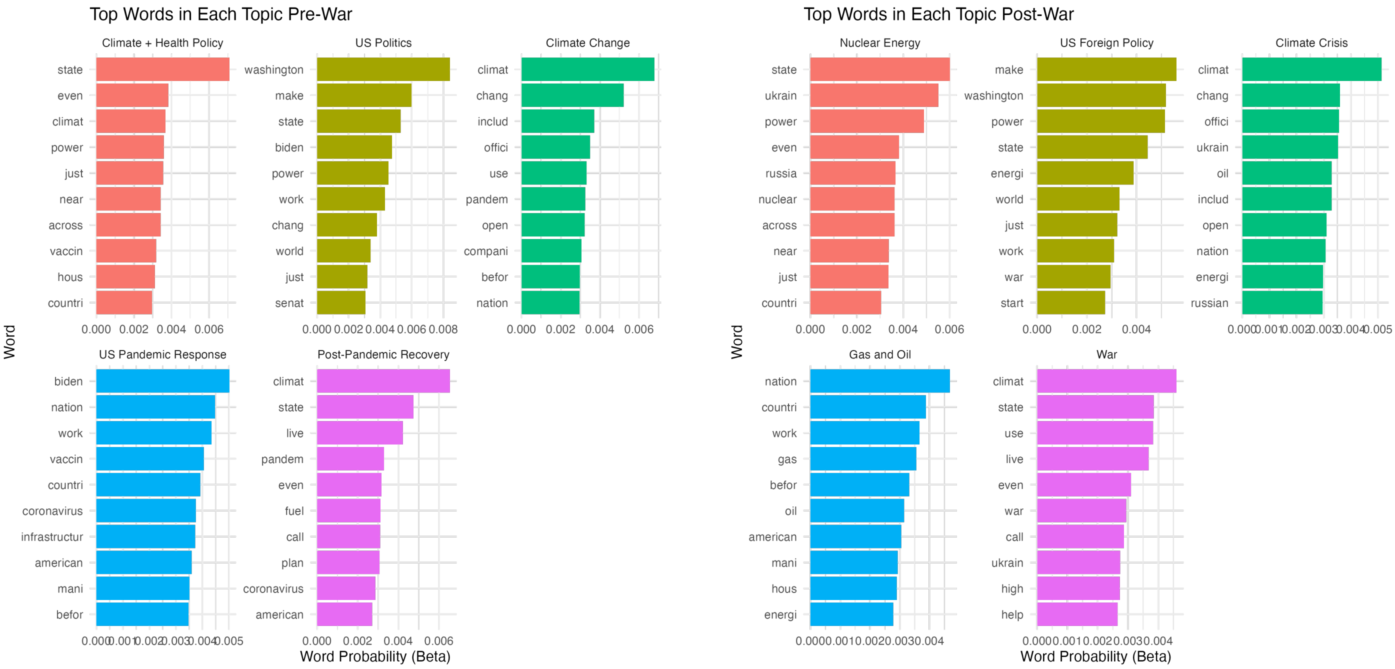
Table 1. Regression Results for Model 1

The results from this initial model are inconclusive: the coefficient for the war variable is not statistically significant and does not exhibit the expected positive sign.

Next, I conduct a second estimation to examine whether the impact of the war on energy salience evolved over time. I extend the model by incorporating a continuous variable that captures the number of days relative to the start of the conflict—negative for articles published before February 24, 2022, and positive for those published after. This allows me to assess whether attention to energy topics increased, decreased, or remained stable in the aftermath of the conflict, offering a more nuanced analysis than the simple pre/post-war comparison. I also include an interaction term between the war indicator and the time-elapsed variable to allow for distinct time trends before and after the conflict. This specification helps determine whether any observed rise in energy salience was sustained over time or merely reflected a short-term reaction to the war.

[1] Coherence measures the semantic similarity of top words in each topic, with higher scores signifying more interpretable topics. The first local maximum in coherence (ee Appendix A.1. Figure 6) occurs at K = 5. Perplexity, which gauges how well the model predicts unseen data, reaches its minimum around K = 22 (see Appendix A.1. Figure 7). While the 22-topic model fits better statistically, K = 5 offers a balance between interpretability and complexity.

[2] The LDA topic modeling results disaggregated by year reveal a shift in media discourse: in 2021, themes focused on COVID-19 and U.S. politics, while 2022 saw a shift toward energy security, rising gas and oil prices, nuclear energy, and the Russia-Ukraine war. By 2023, energy topics declined, appearing in only one topic, indicating reduced media focus as the immediate geopolitical and economic impacts of the war stabilized (see Appendix A.2. for details).



(a) Pre-War **(b)** Post-War

Figure 3. Topic Modeling using Latent-Dirichet Allocation

$$EnergyScore = \beta_0 + \beta_1 War + \beta_2 TimeTrend + \beta_3 Time * War \quad (2)$$

Variable	Energy Score
Intercept	3.334***
	(0.0832)
War	0.0009
	(0.1052)
Time Trend	0.0006
	(0.0003)
War*TimeTrend	-0.0014***
	(0.0004)
F(3,9295)= 5.162 p-value: 0.0015	

Table 2. Regression Results for Model 2

The war coefficient is not statistically significant in either Model 1 or Model 2. However, the interaction term in Model 2 indicates that the effect of the war on energy salience diminishes over time. This trend aligns with earlier findings: while media attention to energy policy spiked sharply in 2022 following the outbreak of the war, it notably declined by 2023. As a final step, I conduct a Chow Test to assess whether a structural break occurred at the onset of the war. The test yields a statistic significant at the 1% level, leading to the rejection of the null hypothesis of no structural break. This result provides strong evidence of a significant change in how energy-related content was framed in news coverage following the start of the Russia–Ukraine conflict. Specifically, the relationship between time and the energy salience score shifts markedly at this point, indicating a structural break in the data.

Chow Test	Statistic	P-value
	6.005	0.0025
Reject H_0 of no structural break		

4.4. Structural Topic Modeling

Finally, I apply a Structural Topic Model [5] to analyze how energy-related news coverage evolved from 2021 to 2023, particularly in response to the Russia–Ukraine war. STM incorporates document-level covariates—such as time, a binary war indicator, and their interaction—enabling a more nuanced analysis of how topic prevalence varies across contexts and over time.

I estimate a five-topic STM, consistent with the earlier LDA model. To capture complex, non-linear trends in topic prevalence, I include regression splines with five degrees of freedom. This allows the model to flexibly capture complex patterns without strict functional assumptions.

By controlling for both the war's onset and broader temporal dynamics, this framework provides insight into whether specific topics became more or less prominent over time—and how the war may have shaped the trajectory of energy-related discourse.

The results of the five-topic STM are presented in Table 3. I assigned descriptive labels to each topic based on the highest-probability and FREX words.[3]

Next, I analyze how the prevalence of each topic changed after the war (see Figure 4). Positive values indicate an increase in topic prevalence post-war, while negative values indicate a decline.

Topics 1 and 4—centered on greenhouse gas emissions and the energy crisis, respectively—become more prevalent after the onset of the war, reflecting the increased attention to energy security and geopolitical tensions. In contrast, Topic 5, which focuses on gas prices, declines in prevalence. While this may appear counterintuitive, it likely reflects a shift in media framing—from specific price fluctuations to broader narratives around global energy markets and security. This trend may also be driven by a general decline in public and media attention to gas prices over time.

Topic 2, covering U.S. politics, shows a decline in prevalence following the war, whereas Topic 3, related to lifestyle content, experiences a slight increase. However, the magnitude of these changes is relatively limited, suggesting that although geopolitical shocks can influence media coverage, longer-term structural trends may play a more domi-

[3] FREX attempts to find the words that are both frequent and exclusive, identifying words that distinguish topics.

Table 3. STM Topics with Highest Probability and FREX Words

Topic #	Label	Highest Probability Words	FREX Words
1	Greenhouse Emissions	energi, countri, global, mani, russian, work, war	carbon, energi, inflat, russian, conflict, greenhous, emiss
2	US Politics	climat, presid, unit, washington, nuclear, nation, biden	biden', republican, presid, democrat, tax, washington, polici
3	Lifestyle	time, home, can, wednesday, hurrican, tuesday, govern	room, grill, recip, meat, ice, ingredi, music
4	Energy Crisis	year, one, chang, oil, day, ukrain, will	oil, fuel, will, day, world', crisi, one
5	Gas Price	state, said, gas, peopl, price, thursday, month	price, peopl, electr, rain, high, area, thursday

Note: Topics extracted using STM; Highest Probability and FREX words help interpret topic uniqueness and frequency.

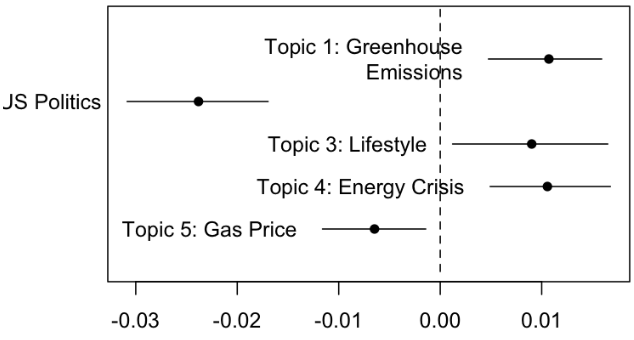


Figure 4. War Effect on Topics

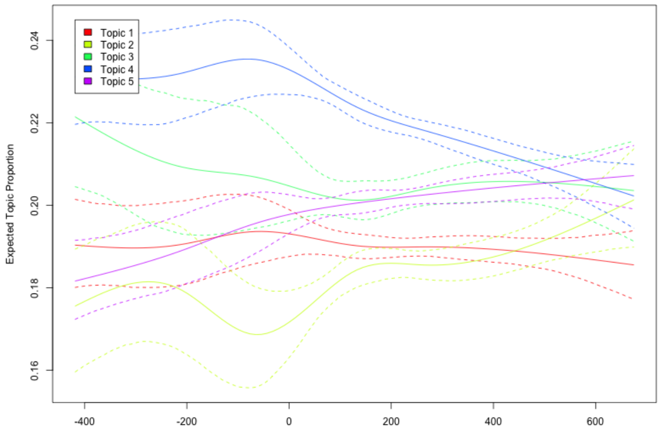


Figure 5. Topics Prevalence Over Time

nant role.

To further examine these dynamics, I use spline regressions to model the temporal effect on topic prevalence (see Figure **??**). Negative values on the x-axis represent days before the war, while positive values correspond to days after its onset.

Several topics exhibit distinct shifts around the beginning of the conflict. Topic 1 (Greenhouse Emissions) receives increasing attention in the lead-up to the war but gradually declines as the conflict continues. Topic 4 (Energy Crisis) spikes sharply just before the war, likely reflecting anticipatory coverage, before tapering off over time. In contrast, Topic 5 (Gas Prices) shows a steady increase both before and after the war, possibly driven by sustained economic repercussions. Topic 2 (U.S. Politics) declines sharply during the war, indicating a shift in media focus away from domestic political issues toward energy and international affairs. Finally, Topic 3 (Lifestyle) experiences a temporary drop in prevalence during the war, but its coverage soon returns to prior levels.

Overall, these findings are consistent with the exploratory analysis presented earlier: in 2022, top words were heavily centered on energy-related themes, reflecting heightened concern during the onset of the war. However, by 2023, this emphasis had diminished. As the spline regressions show, energy-related topics surged around the start of the conflict but gradually normalized over time, suggesting a reversion to broader media narratives.

5. Discussion

The analysis of shifts in energy policy discourse before and after the Russia-Ukraine war reveals several key insights. First, the surge in negative sentiment following the onset of the conflict highlights how geopolitical crises can swiftly influence public perception and media narratives surrounding energy policy. LDA topic modeling shows a clear shift in focus—from broader issues like climate change, pandemic politics, and economic concerns to an increased emphasis on energy security and geopolitics. This shift signals a recalibration of priorities in response to the global crisis.

The gradual decline in the salience of energy issues, as observed in the interrupted time series analysis, suggests that while the war initially intensified media coverage, the urgency surrounding energy policy diminished over time as other global concerns came to the forefront. This pattern underscores the fleeting nature of media attention, particularly for issues like energy policy, which often recede once the immediate crisis subsides.

Structural topic modeling further highlights the evolving salience of different energy-related topics, showing how some gained prominence during the war, while others surged temporarily only to fade as the immediate crisis subsided.

In conclusion, the analysis illustrates how a major geopolitical event like the Russia-Ukraine war can significantly reshape the discourse surrounding energy policy, steering the conversation toward security and geopolitical concerns. However, this focus, while crucial during the conflict, faded once the immediate crisis waned. This is concerning, given the ongoing climate change crisis and the uncertainty surrounding energy security in the years ahead. Energy policy, being critical for the development of nations, should remain at the forefront of discussions, with a focus on long-term energy security as a priority in today's world.

For future research, I plan to incorporate word embeddings to explore the semantic relationships between key terms in the energy policy discourse. This method allows for a more nuanced analysis of how concepts such as nuclear power, renewable energy, and sustainability are framed and interrelated in media coverage. By mapping these associations, word embeddings can help uncover underlying narratives and policy framings that may not be fully captured through topic modeling alone.

A. Appendix

A.1. Metrics to choose optimal K

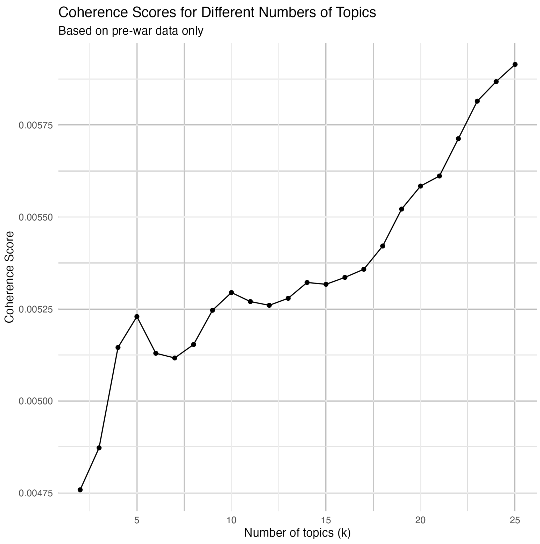


Figure 6. Coherence Score to choose optimal K

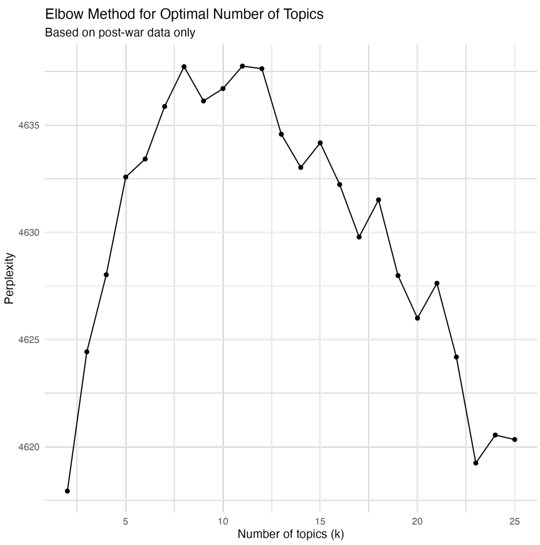


Figure 7. Perplexity Score to choose optimal K

A.2. LDA Topic Modeling by year

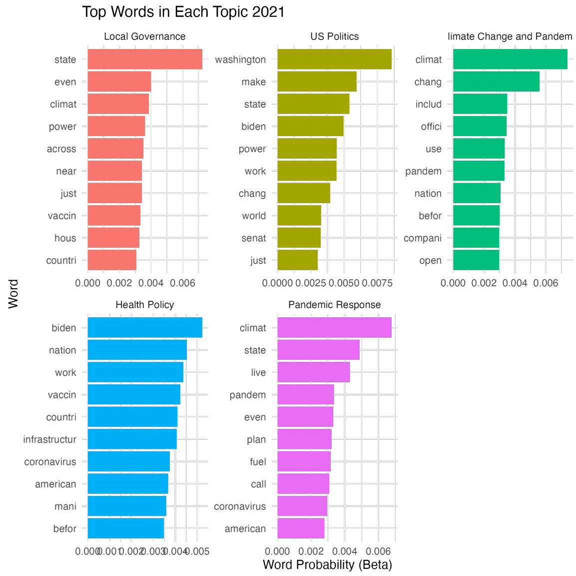


Figure 8. Topic distribution, 2021

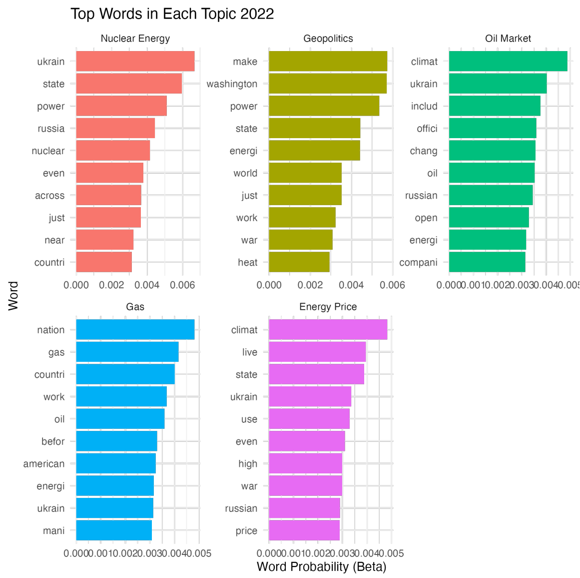


Figure 9. Topic distribution, 2022

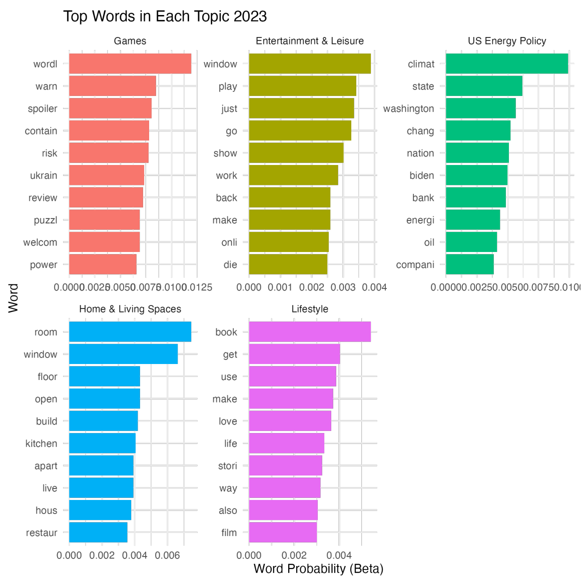


Figure 10. Topic distribution, 2023

References

[1] M. Balmaceda, P. Högselius, C. Johnson, H. Pleines, D. Rogers, and V.-P. Tynkkynen, *Rethinking energy materialities in the shadow of Russia's war on Ukraine*, Energy Research Social Science, 2024. [Online]. Available: https://doi.org/10.1016/j.erss.2024.103678.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, *Latent Dirichlet Allocation*, Journal of Machine Learning Research, 2003. [Online]. Available: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

[3] P. L. Bøggild and J. E. Thorsen, *Topic Modelling Three Decades of Climate Change News in Denmark*, SSRN, 2023. [Online]. Available: https://ssrn.com/abstract=4513921.

[4] V. A. Nieves and J. Meckling, *A historical turning point? Early evidence on how the Russia-Ukraine war changes public support for clean energy policies*, Energy Research Social Science, 2022. [Online]. Available: https://doi.org/10.1016/j.erss.2022.102758.

[5] B. M. Stewart, D. Tingley, and M. E. Roberts, *R Package for Structural Topic Models*, CRAN, 2019. [Online]. Available: https://cran.r-project.org/web/packages/stm/vignettes/stmVignette.pdf.