Diplomatic Echoes: A Textual Analysis of White House and European Commission Communications on Ukraine

Rui Maciel, Viktoriia Yuzkiv

March 15, 2024

Abstract

This paper conducts an analysis of the official communications from the leaders of the White House and the European Commission to quantify and understand the impact of the Russia's war against Ukraine on their respective communication strategies. By examining mentions of Ukraine, employing Latent Dirichlet Allocation (LDA) for topic analysis, and applying Granger Causality Tests, we reveal distinct trends in how each entity discusses the war. Our findings indicate that the European Commission tends to follow the White House's lead in terms of the frequency and context in which Ukraine is mentioned. This comparative study hopes to contribute to the literature of diplomatic communication during an international crisis, providing insights into the coordination and emphasis placed on Ukraine by its two allies.

Contents

1 Introduction **2**

2 Literature Review **2**

3 Data Gathering and Pre-processing **3**
3.1 Data Gathering 4
3.2 Pre-processing Steps 4
3.3 Final Datasets 4

4 Time-line of the Coverage **5**
4.1 General Trends in Speeches, Statements, and Press Releases 5
4.2 Share of Documents Mentioning Ukraine 5

5 LDA Topic Analysis **8**
5.1 Methodology and Parameters 8
5.2 Contribution of the White House and the European Commission to the Joint Dataset 10
5.3 Time analysis of the topic of interest 12

6 Granger Causality Test **13**
6.1 Coverage data 14
6.2 Topic shares data 14

7 Conclusion and Next Steps **15**

A Trends in the Number of Speeches, Statements, and Press Releases (The White House and The European Commission) **17**

B Top Words Associated with Each Topic **18**

1 Introduction

On February 24, 2022, an event reminiscent of Cold War tensions unfolded. On this day, Russia, the successor state of the Soviet Union, initiated a full-scale invasion of Ukraine, following its illegal annexation of a region within one of its former republics. This act not only revived the specters of the Cold War but also drew immediate and widespread reactions from NATO, the United States, and European leaders. Despite the significant attention the Ukrainian struggle garnered initially, a comprehensive quantification of the significance and coverage of the Russo-Ukrainian war by the United States and the European Union has yet to be conducted.

This paper seeks to address this gap by quantifying coverage trends and investigating the events that may have led to increased attention towards Ukraine. We aim to assess the impact of such coverage on international diplomacy and policy-making.

To achieve our objectives, we do an analysis of official communications from the White House and the European Commission that were made by its respective leaders, President Joe Biden and the President of the European Commission Ursula Von der Leyen. Our preliminary work involves preprocessing to discern general trends in documents mentioning Ukraine. We then employ Latent Dirichlet Allocation (LDA) analysis, to explore the predominant topics within these communications. Our goal is to map the evolution of discourse related to Ukraine and its significance on its biggest allies.

Our analysis begins with a literature review, followed by an examination of the data gathering, preprocessing, and the construction of final datasets. We then analyze the coverage timeline, delve into LDA topic analysis to understand the thematic focus, and employ Granger Causality Tests to investigate causative relationships in communication strategies.

Our approach reveals that the European Commission's communication strategy on Ukraine often follows the lead set by the White House.

2 Literature Review

In the realm of contemporary communication, there exists a growing interest in employing innovative methodologies to decipher the nuances of textual content, particularly in political discourse. Within this context, techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Dirichlet Allocation (LDA) analysis have emerged as the basic tools for topic exploration [1, 2].

For instance, recent studies have endeavored to scrutinize the linguistic patterns present in the speeches of prominent leaders such as Von der Leyen, Biden, and Jens Stolberg, utilizing LDA analysis to discern prevailing themes. Notably, their analysis revealed a predominant focus on diplomatic resolutions and the prevention of escalation during the critical periods of February 2022 and December 2022 [7].

Furthermore, the invasion of Ukraine sparked a surge in media coverage analyses, aimed at dissecting the narratives and linguistic choices evident in regional reporting. Through the application of LDA analysis, researchers observed stark disparities between Western, Russian, and Chinese media outlets, particularly in their terminology regarding the conflict. While Western media readily adopted the term "invasion," Russian sources utilized euphemistic language like "operation," and Chinese media opted for the neutral term "conflict," highlighting the influence of political and geographical contexts on media discourse [5].

Moreover, scholarly investigations have shed light on Russia's concerted efforts to propagate its version of events, which have met formidable opposition from Western media outlets seeking to counteract and clarify the misinformation disseminated by Russian channels [6]. Employing advanced techniques such as Neural Topic Modeling coupled with class-based TF-IDF procedures, researchers have unveiled the intricate dynamics at play within the information landscape.

Additionally, the Granger Causality test, commonly used in comparative analysis across different regions, offers a robust method to explore potential causal relationships between the communications of various entities. For instance, researchers utilized the Granger Causality test to investigate the influence of news media and public opinion on the communication strategies of international organizations [3]. By incorporating the Granger Causality test into our study of the White House and European Commission communications regarding Ukraine, we aim to unveil potential causal relationships between these two prominent entities, thereby enriching our understanding of the complex dynamics surrounding the Russo-Ukrainian war.

This convergence of media analysis and technological methodologies underscores the multifaceted nature of modern conflict. As this paper embarks on an analysis of official communications from the White House and European Commission leaders regarding Ukraine, it does so with the aim of leveraging LDA topic analysis to elucidate underlying trends and dynamics. By applying these methodologies, our study seeks to fill the gap in understanding the patterns and nuances within official communications, providing valuable insights into the international response to the Russo-Ukrainian war.

3 Data Gathering and Pre-processing

In this chapter, we outline the methodologies employed for data collection and preparation from two distinct sources — the White House Briefing Room and the European Commission Press Corner, covering the period from January 2021 to December 2023. Notably, for the European Commission dataset, our focus was specifically directed towards communications attributed to Ursula von der Leyen, the President of the European Commission during the study period, and for the White House one - exclusively to those attributed to President Biden. This deliberate selection aligns with our objective of comparing and contrasting the leadership communications strategies of Ursula von der Leyen and President Biden.

3.1 Data Gathering

To acquire the necessary data, we developed our own scraping pipeline, encompassing the period from January 2021 to December 2023:

- **White House Data:** The data was retrieved from the publicly accessible website White House Briefing Room [8]. We gathered documents falling under various categories, including 'Speeches and Remarks', 'Statement releases', and 'Press briefings'.
- **European Commission Data:** Similarly, the data for the European Commission was accessed via its official website European Commission Press Corner [4]. Our focus was on documents classified under 'Statements', 'Speeches', and 'Press releases'.

3.2 Pre-processing Steps

After receiving the raw data, we proceeded with a comprehensive pre-processing pipeline aimed at refining and standardizing the datasets. The key pre-processing steps undertaken are detailed below:

- **Standardizing Date Formats:** To streamline analysis, the datasets were filtered to include records spanning the years 2021 to 2023. This delineation facilitated the creation of a comprehensive three-year panel dataset, with the inclusion of 2021 allowing for comparisons with the period before the full-scale invasion.
- **Updating Category Names:** Category labels were harmonized across both datasets, ensuring consistency and comparability in subsequent analyses.
- **Exclusion of Non-English Texts (European Commission Dataset):** Non-English texts were identified and subsequently excluded from the European Commission dataset to maintain data integrity and mitigate language-related biases.
- **Lemmatizing Data:** Textual data underwent lemmatization — a linguistic process aimed at reducing words to their base or dictionary form. This standardization enhanced the comparability and interpretability of textual content, bolstering the analytical capabilities of the datasets.
- **Adding Binary Column for Mentions of Ukraine:** A binary column was introduced to detect mentions of Ukraine within the datasets, providing a valuable indicator for tracking relevant instances and enabling focused analysis. We consider that a document mentions Ukraine if it contains at least one instance of the words 'Ukraine' or 'Ukrainian'.

3.3 Final Datasets

The culmination of the pre-processing efforts yielded two refined datasets, each primed for detailed analysis:

- **The White House dataset** comprises 3855 documents, meticulously curated to ensure data integrity and consistency.

- **The European Commission dataset**, post-pre-processing, comprises 1078 documents, encompassing a diverse array of communications disseminated by the Commission.

4 Time-line of the Coverage

4.1 General Trends in Speeches, Statements, and Press Releases

Before delving into the share of mentions of Ukraine, it's essential to understand the distribution of documents in our datasets. This section provides insights into the general trends observed in speeches, statements, and press releases from both the European Commission and the White House.

Communications attributed to Ursula von der Leyen in the European Commission demonstrate a relatively consistent level of communication across its three main categories: speeches, statements, and press releases. On average, there are approximately 8 speeches, 12 statements, and 11 press releases per month. This indicates that the Commission maintains a balanced approach in its communication strategies, with similar levels of activity across these categories.

At the same time, communications attributed to the President from the White House demonstrate notably higher levels — ranging from 2 to 4.3 times higher — compared to those emanating from the European Commission. Similarly to the Commission's trends, statements represent the most prevalent category, with an average of 52 statements per month. However, the White House surpasses the European Commission in both speeches and press releases, with averages of 34 and 22 per month respectively.

The time-lines of the number of speeches, Statements, and Press Releases for both the European Commission and the White House are provided in the Appendix A.

4.2 Share of Documents Mentioning Ukraine

The share of mentions of Ukraine offers valuable insights into the attention and focus given to Ukraine in the communications of both the European Commission and the White House. A mention of Ukraine is defined as at least one occurrence of "Ukraine" or "Ukrainian" in the document.

Both the European Commission's and the White House's communication regarding Ukraine exhibit fluctuations in the share of mentions over time. Before the full-scale invasion, there were barely any mentions of Ukraine in either. Notably, there is an observable increase in mentions starting from January 2022, preceding the full-scale invasion that occurred on February 24, 2022, as that is when discussions on whether Russia would invade Ukraine began.

The share of mentions of Ukraine by the European Commission reached its peak of 84% during the first month following the invasion. Subsequently, there has been a gradual decline in mentions over time [Figure 1], although occasional spikes can still be observed. These spikes can be attributed to various factors, including major events directly related to the war, discussions concerning financial aid for Ukraine or the refugee situation, as well as visits by Volodymyr Zelenskyy to Europe or the US, along with visits by European or US representatives to Ukraine.

Similarly, the White House's communication regarding Ukraine also demonstrates fluctuations in the share of mentions over time, albeit at a significantly lower level compared to the European Commission's. The highest share was observed in March 2022, reaching a level of 60%. Over

the past year, the share of documents mentioning Ukraine has generally ranged between 21% and 38%.

Figure 1: Share of Documents Mentioning Ukraine by Month

It is also worth examining the shares by document type to uncover interesting observations.

For the European Commission [Figure 2], it is noteworthy that in April 2022, January and February 2023, and July 2023, 100% of speeches mentioned Ukraine. While all three categories display a relatively high level of mentions of Ukraine, it is observed that the share is lowest in press releases, and the decreasing trend is most noticeable in this category.

For the White House [Figure 3], the largest share of mentions of Ukraine is observed in press releases, contrary to the trend observed in the European Commission's communication. Notably, the share of mentions in press releases remains consistently high, averaging around 81% for the years 2022-2023. Conversely, the share of mentions in speeches and statements exhibits a decreasing trend over time, with speeches generally having slightly higher coverage overall compared to statements.

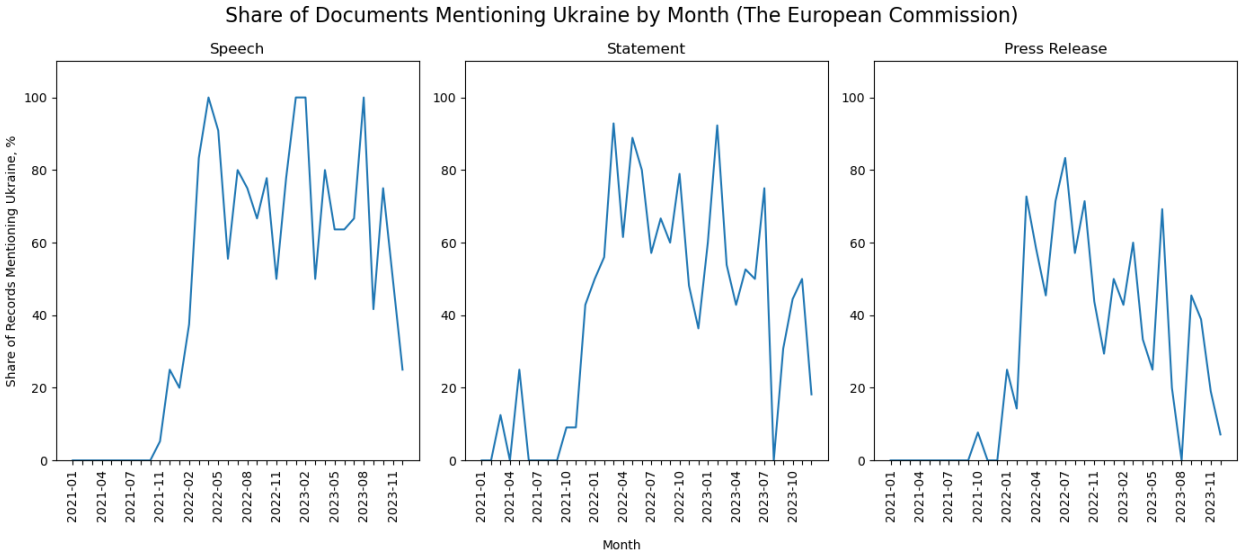

Figure 2: Share of Documents Mentioning Ukraine by Month (The European Commission)

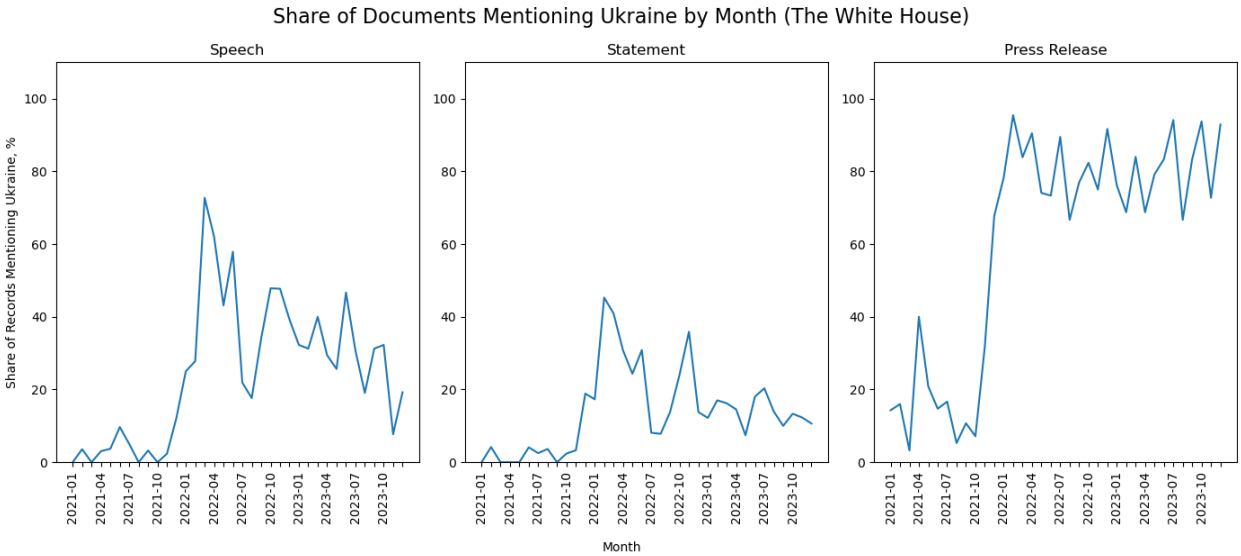

Figure 3: Share of Documents Mentioning Ukraine by Month (The White House)

5 LDA Topic Analysis

5.1 Methodology and Parameters

In this study, we utilized Latent Dirichlet Allocation (LDA) to identify topics within official communications from the European Commission and the White House.

LDA, a probabilistic model, groups words into topics based on their distribution across documents, capturing the contextual relationships between them. To enhance the reliability of our analysis, we complemented LDA with Count Vectorization and TF-IDF (Term Frequency-Inverse Document Frequency).

Count Vectorization tallies word frequencies within documents, while TF-IDF assesses a term's uniqueness across the corpus, distinguishing between common and meaningful terms. By combining these techniques, we aimed to leverage the strengths of both approaches.

We refined our analysis through parameter tuning of the LDA, Count Vectorization, and TF-IDF models. This involved adjusting the number of topics for LDA and setting term frequency thresholds for TF-IDF to minimize the impact of irrelevant words and outliers.

In order to be able to compare directly the European Commission data with the White House official data, the LDA model was run in a joint dataset were both were present.

The parameters tested encompassed all combinations of 0.005, 0.01, 0.05, 0.1 for minimum document frequency, and 0.2, 0.3, 0.4, 0.5, 0.6 for maximum document frequency. The chosen combination of topics was the one that appeared most logical to the human eye and had the least noise. In this context, "noise" is defined as a combination of words that do not convey a clear topic and consist of vague verbs, pronouns, or vocabulary used for communication purposes. To see all of these combinations please see the file attached to this paper.

The final parameters used were 0.1 for minimum document frequency and 0.5 in maximum document frequency. To find the results for CV and TF-IDF, please consult appendix B. Below are the LDA topics:

Results of LDA
Topic 1: security, US, global, partnership, pacific, economic, technology, national, cooperation, commitment
Topic 2: serve, national, university, member, director, US, house, secretary, department, law
Topic 3: right, woman, thank, vice president, nation, vice, fight, many, leader, family
Topic 4: go, know, get, say, think, thank, want, look, thing, time
Topic 5: community, administration, program, health, federal, provide, million, department, access, service
Topic 6: ukraine, russia, war, minister, security, prime, prime minister, president biden, military, russian
Topic 7: vaccine, get, health, thank, day, COVID19, question, know, pandemic, week
Topic 8: go, say, know, see, would, think, get, question, right, want
Topic 9: energy, climate, european, investment, EU, clean, global, resilience, europe, commission
Topic 10: job, cost, pay, american, investment, america, million, economy, worker, infrastructure

Table 1: Top words associated with each topic (LDA)

Upon examining the themes generated via Latent Dirichlet Allocation (LDA), we observe a distinction between topics focused on American concerns and those related to European interests. This difference is illustrated in Topic 9, which covers European policy on energy and climate within its resilience framework, and Topic 10, which addresses the American employment and labor market.

An important observation within this analysis is Topic 6's focus on the war in Ukraine, showing shared themes across both datasets with terms such as 'Ukraine,' 'Russia,' and 'war.' The mention of President Biden introduces an American-centric viewpoint. This prompts further investigation to understand each dataset's contributions to these topics. Before proceeding, let's consider the specifics of each enumerated topic:

- **Topic 1** discusses security and global partnerships, possibly reflecting transatlantic defense dialogues and cooperative security efforts. The presence of 'US' and 'pacific' suggests a significant contribution from the White House dataset.
- **Topic 2** concerns academia and governance, potentially related to educational policy and legislative processes within institutional frameworks.
- **Topic 3** deals with societal and civil liberties, touching on gender equity and socio-cultural dynamics. The mention of Kamala Harris, the first vice president of the United States, correlates with this topic.
- **Topic 4 and 8** provide limited insights into policy discussions or thematic concerns, highlighting the challenges in linguistic data analysis and the need for contextual understanding in topic modeling. These will be excluded for the next visualizations.

- **Topic 5** focuses on public health and government programs, reflecting on healthcare provision and the role of federal agencies. This indicates a more US-centric approach, given the ongoing political debate in the US regarding public healthcare expansion.
- **Topic 7** concentrates on the healthcare sector, emphasizing vaccination efforts against COVID-19. This aligns with the dataset including data from the beginning of 2021, when vaccine rollout occurred in both continents.

5.2 Contribution of the White House and the European Commission to the Joint Dataset

Following the examination of individual topics, we advance our analysis by quantitatively assessing the thematic contributions from each dataset, specifically those from the European Commission (EC) and the White House (WH), to the overarching topic distribution. This step is crucial for quantifying each dataset's influence within the combined corpus, providing a statistical foundation to support our qualitative assessments.

To achieve this, we analyze the topic distributions derived from each dataset. This involves calculating the frequency with which each topic appears across all documents within a dataset, determining the average prominence of each topic. For both datasets, we ensure that every topic is considered, regardless of its probability, to make sure we get a full picture. This method allows us to compile an average distribution for topics across the documents, shedding light on the dominant themes characterizing each corpus.

By comparing these average distributions, we gain insights into the thematic emphases of the EC and WH datasets within the joint analysis. This comparison enables us to quantitatively assess the contributions of each dataset to the overall topic model. It not only verifies whether our earlier observations were accurate but also clarifies the extent to which each corpus influences the combined topic landscape. Through this analytical approach, we develop a structured method to quantitatively ascertain the specific thematic contributions of the European Commission and White House datasets, offering a detailed understanding of their respective impacts on the collective topic distribution.

We have labeled the topics as follows to enhance readability in the upcoming charts: 1) Economic Diplomacy, 2) Government and Public Service, 3) Social Rights and Leadership, 5) Health and Community Services, 6) International Conflict and Ukraine, 7) Public Health and COVID-19, 9) European Energy and Climate Policy, 10) American Economic Development and Labor.

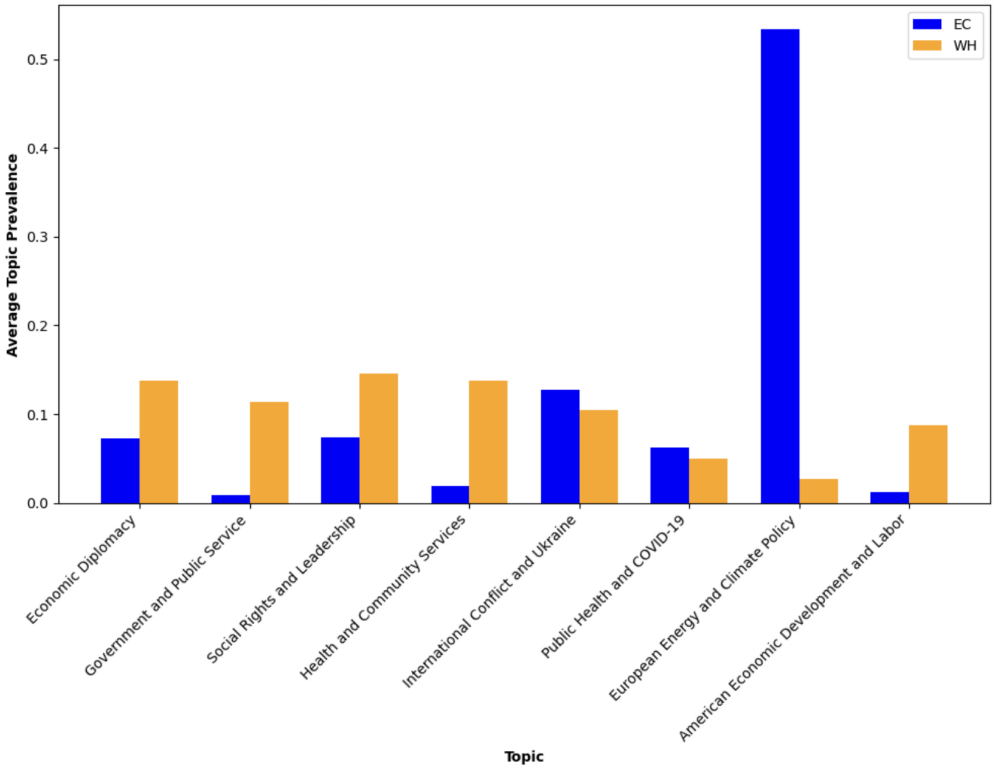

Figure 4: Average Topic Distribution Comparison between the European Commission (EC) and White House (WH)

The analysis of the topic distribution reveals a slight edge of the United States in most thematic areas, with exceptions in discussions surrounding the Russo-Ukrainian war, the COVID-19 pandemic, and, quite naturally, European Energy and Climate Policy. The latter topic's pronounced representation within the European Commission (EC) dataset validates the Latent Dirichlet Allocation's (LDA) capability to effectively differentiate and classify topics, reflecting a natural propensity for European-centric policy terms to populate the EC's communications.

Concurrently, the 'American Economic Development and Labor' topic exhibits a bias towards the White House (WH) dataset. This preponderance likely mirrors the inward focus of U.S. policy discourse on national economic and labor issues, underscoring the topic's relevance and alignment with American administrative communication.

The analysis of 'Economic Diplomacy' presents a nuanced picture. While terms such as 'US' and 'pacific' might initially suggest an American orientation, the inclusion of terms synonymous with collaborative endeavors indicates meaningful European engagement within this sphere. This interpretation is further supported by the European Union's diplomatic activities, including significant trade agreements with Pacific entities, such as the landmark EU-New Zealand trade accord established in 2022.

In the examination of the 'International Conflict and Ukraine' topic, we observe one of the more equitably represented subjects across the datasets, albeit with a slight elevation in the EC's data. This balanced distribution implies the topic's significant resonance within the discursive fabric of both the European and American dialogues, transcending geographical and jurisdictional

boundaries to reflect a topic of global pertinence and shared concern.

5.3 Time analysis of the topic of interest

Turning our attention into the main topic of interest 'International Conflict and Ukraine', we observe its variations within the aggregate dataset, which uses data from both the European Commission (EC) and the White House (WH).

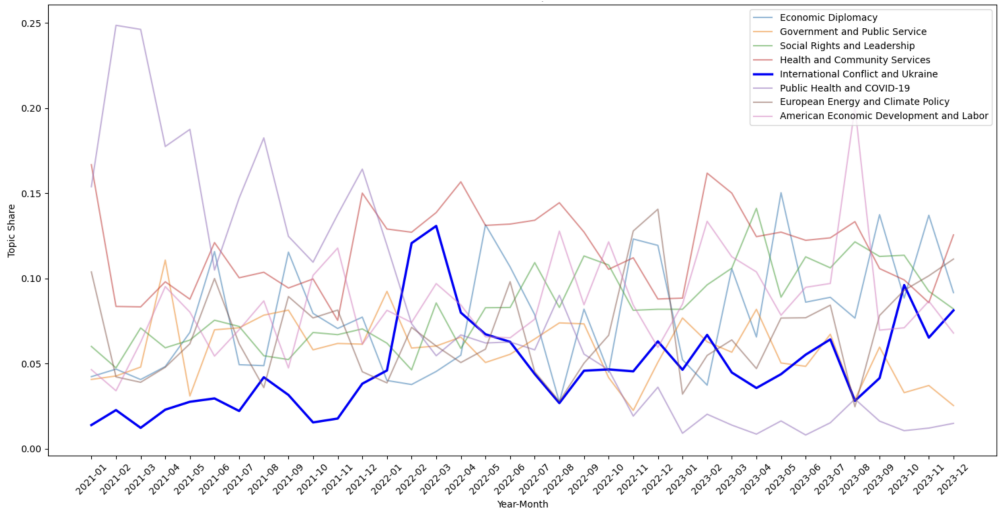

Figure 5: Trend of Topic Shares in the Joint Dataset

Focusing on the issue of international conflict, particularly in the context of Ukraine, February 2024 marks a significant juncture with the Russian invasion. This notable spike in data underscores the model's adeptness at capturing pivotal global events. Another noteworthy peak in October 2023 suggests that while the situation in Ukraine commands attention periodically, domestic concerns within the U.S. and the EU maintain considerable prominence within the datasets. For instance, discussions revolving around health matters in the U.S. and energy policies in the EU during December 2022 exhibit robust significance. Thus, it becomes evident that local issues wield substantial influence on the broader discourse, even amidst pressing global affairs such as those in Ukraine.

The initial salience of the 'Public Health and COVID-19' theme within the dataset reflects the model's efficacy in capturing prevailing discourses. Its early prominence aligns closely with the tangible escalation of the pandemic, thereby affirming the model's proficiency in tracking lexical trends in sync with real-world developments. This synchrony between the outputs of the LDA model and the chronological unfolding of events serves to validate its effectiveness in delineating the lexical landscape vis-à-vis real-world occurrences.

As now we understand both the contribution and trends to the topic of our interest, we will turn our analysis to uncover potential lead-lag patterns among various topic areas, especially in relation to Ukraine.

6 Granger Causality Test

In this section, we delve into the application of the Granger Causality Test to explore potential causal relationships between communications in the European Commission and the White House. This analysis seeks to provide insights into the dynamics of communication strategies employed by these entities and their potential influence on each other regarding discussions related to Ukraine. By understanding the direction of causality between these variables, we aim to uncover valuable insights into the relationship between the European Commission and the White House in terms of their communication regarding Ukraine during the years 2021-2023.

The Granger Causality Test is a statistical hypothesis test used to determine whether one time series variable can predict another. Developed by Nobel laureate Clive Granger, this test is widely used in econometrics and time series analysis to infer causality between variables. The fundamental idea behind the Granger Causality Test is that if a variable X "Granger-causes" another variable Y, then past values of X should contain information that helps predict current values of Y, beyond what can already be predicted from past values of Y alone.

In our analysis, we apply the Granger Causality Test to investigate whether mentions of Ukraine in the European Commission's communications can predict or influence mentions of Ukraine in the White House's communications, and vice versa. Additionally, we will incorporate data on the Ukraine-related topic discussed in both the White House and the European Commission to provide a comprehensive analysis of communication dynamics.

To conduct the Granger Causality Test, the data is aggregated by week. The decision to aggregate the data on a weekly basis stems from the recognition of significant fluctuations and noise observed in daily data. These fluctuations often arise from irregular publishing schedules, random events, or short-term variations in public attention. Aggregating the data weekly helps mitigate these fluctuations, resulting in a more stable and interpretable trend over time.

Before proceeding with the Granger Causality Test, it is crucial to assess the stationarity of the time series variables. Stationarity, a key assumption for the test, implies that the statistical properties of the series remain constant over time. This is typically verified using statistical tests such as the Augmented Dickey-Fuller (ADF) test. If the time series are found to be non-stationary, differencing is applied to induce stationarity.

After aggregating the data and ensuring its stationarity, the next step involves determining the lag order for the Granger Causality Test. The lag order represents the number of past observations included in the model and is a crucial parameter for the test. In our analysis, we set the maxlag coefficient to 1. This decision is influenced by the characteristics of our data, which exhibit relatively short-term dynamics so that we can focus on assessing causal relationships between the current and immediate past observations, providing insights into more immediate causal dynamics. Additionally, choosing a lower lag order helps mitigate potential overfitting issues and ensures the test's robustness.

The datasets under consideration include the number of documents mentioning Ukraine over time and the share of topics related to Ukraine. Each dataset was tested for stationarity (the ADF test), and differencing was applied where needed. This rigorous preprocessing ensures that the data is suitable for the subsequent Granger Causality Test, facilitating robust and reliable results.

6.1 Coverage data

We first conducted the test on the absolute number of weekly documents mentioning Ukraine to determine if there's a significant effect between the institutions' communications. Specifically, we assessed whether an increase in mentions of Ukraine in one institution predicts a corresponding increase in the other, and vice versa.

	The European Commission_x	The White House_x
The European Commission_y	1.0000	0.018
The White House_y	0.8698	1.000

Figure 6: Granger Causality Test Results for the Mentions Time Series

The p-values of the test are displayed in Figure 6. If the p-value is below 0.05, we should reject the null hypothesis, which would suggest that there is evidence of Granger causality at a 5% significance level. In our case, the p-value for X=The White House and Y=The European Commission is 0.018. The hypotheses for this test are as follows:

- H0 (Null Hypothesis): The White House's lagged values do not significantly affect the prediction of the European Commission communications. In other words, there is no Granger causality between the two variables.
- H1 (Alternative Hypothesis): not H0.

Therefore, we reject the null hypothesis and conclude that the White House will likely be helpful in predicting the European Commission's communication towards Ukraine. On the other hand, changes in the European Commission communications do not Granger cause the White House communication. This means that changes in the European Commission communications do not improve the White House communications prediction performance. This is exactly what we expected. Since communications attributed towards Ukraine are firstly related to geopolitics and warfare, the EU is likely to follow America's communications related to it. For example, Olaf Scholz did not announce the decision to send 14 Leopard 2 tanks to Ukraine until US President Joe Biden's administration hadn't announced plans to send at least 30 M1 Abrams tanks. This test results confirms the narrative that the European Union is dependent on the United States in the field of defense.

6.2 Topic shares data

We conducted an additional test on the weekly time series data from both the White House and the European Commission, focusing on the main topic of interest 'International Conflict and Ukraine'. The results of this test, displayed in Figure 7, mirror those of the number of documents mentioning Ukraine. Notably, the p-value when X=The White House and Y=The European Commission is even lower.

This test's outcomes confirm the prevailing narrative suggesting a dependency of the European Union on the United States in matters concerning defense.

	The European Commission_x	The White House_x
The European Commission_y	1.0000	0.0024
The White House_y	0.6358	1.0000

Figure 7: Granger Causality Test Results for the Topic Time Series

7 Conclusion and Next Steps

To dissect the international narrative and policy shifts surrounding the Russia's war against Ukraine, this paper employs a systematic analysis of the official communications by the leaders of the White House and the European Commission. These sources are pivotal for several reasons. First, they offer direct insights into the official stances and priorities of the United States and the European Union, two major actors in the geopolitical landscape concerning Ukraine. Second, by examining these communications, we can trace the evolution of diplomatic and policy responses over time, offering a timeline of reactions to key events in this war.

We employed Latent Dirichlet Allocation (LDA) for our analysis, aiming to identify the dominant topics across our dataset. Remarkably, this technique proved adept at pinpointing topics that resonate with occurrences, illustrating the distinct contributions of each dataset to the corpus, as well as aligning the temporal progression of topics with events. The precision with which LDA captured the essence of the discourse underscores its effectiveness in reflecting the thematic evolution observed in the communications against the backdrop of ongoing dynamics.

In our exploration, we delve into applying the Granger Causality Test, a statistical measure aimed at discerning potential relationships within the communication strategies we're reviewing. While we approach this test cautiously, acknowledging its limitations and the risk of premature conclusions, our analysis suggests patterns that align with realities. Notably, it seems evident that the United States maintains a leadership influence within NATO, a fact underscored by its significant military budget outlays. The Granger analysis hints at an inferred causality, suggesting that the communicative approaches of the European Commission may be influenced by the United States' stance on NATO and spending. Additionally, another possible reason could be the inherently slower reaction time of the European Commission compared to the US, which could also influence their communication strategies. These two potential causal factors present promising paths for future research.

Looking ahead to it, there are opportunities to refine our approach. Improving our preprocessing methods to filter out topics infused with colloquial language could enhance the thematic clarity of our analysis. Furthermore, transitioning to a neural network framework could provide a more sophisticated mechanism for modeling, potentially surpassing the capabilities of LDA. Additionally, broadening the scope of our Granger Causality analysis to encompass different time frames would strengthen the robustness of our findings, ensuring they are reflective of enduring trends and relationships rather than artifacts of a specific analytical window.

Despite the constraints of our methodology, this paper strives to contribute meaningfully to the academic discourse by offering a grounded examination of the geopolitical narrative surrounding Ukraine. Our findings not only seek to bridge a gap in the literature but also invite further inquiry into the underpinnings of international relations and communication strategies. Furthermore, by

elucidating the communication trends of the European Commission and the White House, we enhance our understanding of the interplay of diplomacy and policy in response to the Russo-Ukraine war, marking a step towards decoding the tapestry of geopolitical communications.

A Trends in the Number of Speeches, Statements, and Press Releases (The White House and The European Commission)

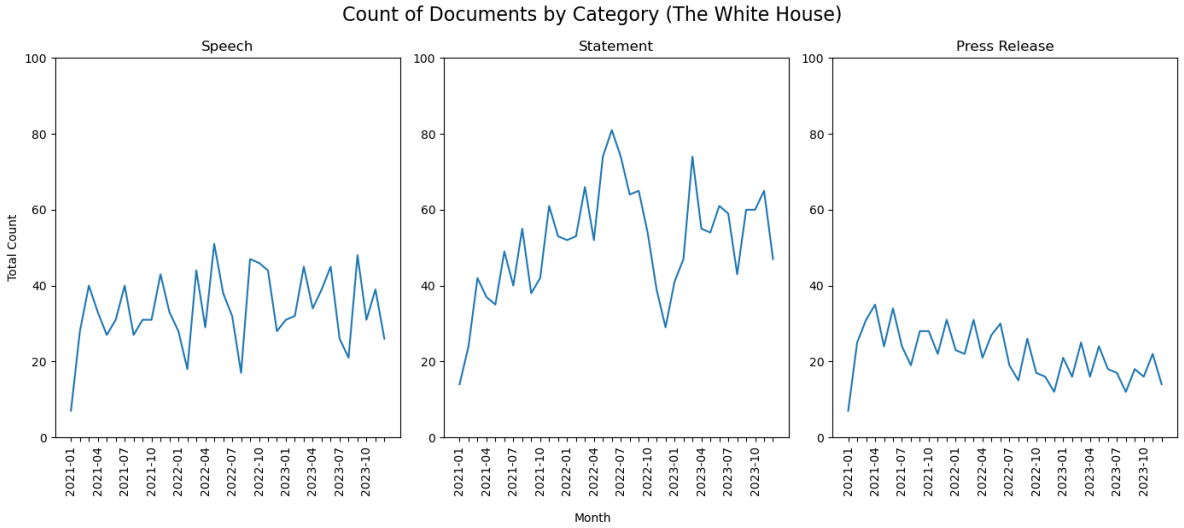

Figure 8: Count of Documents by Category (The White House)

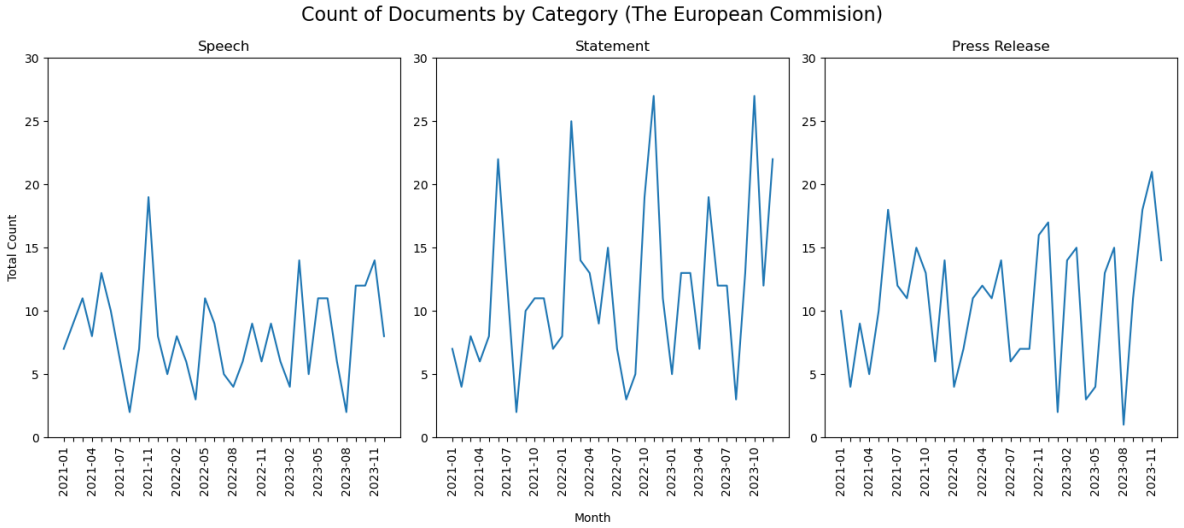

Figure 9: Count of Documents by Category (The European Commission)

B Top Words Associated with Each Topic

Top words associated with each topic (Count Vectorization)
Topic 1: know, say, think, right, want, look, time, question, thank, talk
Topic 2: health, community, program, million, energy, investment, care, climate, provide, access
Topic 3: security, ukraine, global, russia, international, question, ahead, partner, effort, cooperation
Topic 4: energy, climate, world, investment, clean, america, job, economy, global, ukraine
Topic 5: vaccine, thank, health, covid19, dr, global, million, know, world, pandemic
Topic 6: ahead, plan, cost, energy, inflation, price, million, drug, care, reduction
Topic 7: care, health, ukraine, family, child, russia, security, woman, drug, world
Topic 8: vaccine, climate, health, drug, act, inflation, reduction, cost, law, energy
Topic 9: vaccine, ukraine, university, school, serve, member, law, european, price, russia
Topic 10: thank, ukraine, administration, russia, senior, official, drug, cost, think, community

Table 2: Top words associated with each topic (Count Vectorization)

Top words associated with each topic (TF-IDF Vectorization)
Topic 1: go, know, say, get, thank, think, see, want, right, would
Topic 2: security, energy, us, eu, global, national, european, climate, cooperation, discuss
Topic 3: federal, serve, community, health, department, program, university, administration, care, service
Topic 4: european, eu, energy, europe, investment, billion, union, million, plan, clean
Topic 5: federal, ukraine, emergency, assistance, russia, go, question, funding, available, recovery
Topic 6: serve, european, university, eu, ukraine, member, commission, europe, board, district
Topic 7: ukraine, woman, care, family, health, russia, violence, life, child, nation
Topic 8: vaccine, health, covid19, eu, dr, question, european, pandemic, global, vice president
Topic 9: federal, emergency, national, order, thank, joseph, executive order, jr, news, security
Topic 10: vice, vice president, federal, term, harris, district, recovery, emergency, minister, assistance

Table 3: Top words associated with each topic (TF-IDF Vectorization)

References

[1] Akiko Aizawa. An information-theoretic perspective of tf–idf measures. *Information Processing & Management*, 39(1):45–65, 2003.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[3] Omar Dumdum. *Examining the Influence of News Media and Public Opinion in International Institutions: A Political Communication Approach to Studying International Organizations*. PhD thesis, UNIVERSITY OF WISCONSIN–MADISON, 2023.

[4] European Commission. Press corner, February 2024. Accessed: 2024-02-20.

[5] HWA Hanley, D Kumar, and Z Durumeric. "a special operation": A quantitative approach to dissecting and comparing different media ecosystems' coverage of the russo-ukrainian war. *Proceedings of the International ...*, 2023. Cited by 13.

[6] J. Littell and N. Starck. Russian influence operations during the invasion of ukraine. In *International Conference on Cyber Warfare and Security*, volume 18, pages 209–217, February 2023.

[7] Tasente Tanase. Online communication strategy of world political leaders during the Ukraine crisis (February 24 - December 24, 2022). Comparative case study: Joe Biden, Ursula von der Leyen and Jens Stoltenberg. *Technium Social Sciences Journal*, 39(1):207–219, January 2023.

[8] The White House. Briefing room, February 2024. Accessed: 2024-02-20.