Freedom of Press Among World Nations

Shakleen Ishfar Eugene Ayonga

December 14, 2023

1 Introduction

There are many news organizations around the world, who play vital roles in relaying important news of home and abroad to readers. These news often cover incidents that are either positive, negative, or neutral. Moreover, some news stories can be viewed as talking for a country or talking against it.

Freedom of speech is the right to express ideas and opinions without censorship, restraint, or fear of retribution. A news outlet is free if it reports news in an unbiased manner and free from censorship. In this project, we aim to determine whether local news media of a country has freedom of press. To do so, we compare the sentiment and stance of the organization with international news reporting institutions **Reuters** and **Associated Press**. We assume that the more similar a local news is, in terms of sentiment and stance, to the international media outlets, the more free it is.

In summary, this study investigates freedom of speech in local news across countries, examining topic-specific distinctions by comparing sentiment and stance scores with international sources to reveal differences and assess agreement levels.

2 Related Work

While extensive research has explored sentiment classification and scoring within news articles [5, 2], significant efforts have also been dedicated to stance analysis [7, 9]. Furthermore, media freedom and its pillars have been a constant focus for researchers worldwide [1, 11, 10].

However, integrating both sentiment and stance analysis to assess freedom of the press remains relatively underexplored. One notable exception is Rybiński et al.'s (2018) work, which analyzes political news stances to determine press freedom. [12] Additionally, Rybiński (2019) employed machine learning techniques to explore media discourse. [13]

Building upon this existing research, we propose a novel approach that incorporates both sentiment and stance analysis across diverse topics, coupled with statistical inference techniques. This methodology aims to determine whether a national media outlet adheres to the principles of freedom of speech.

3 Methodology

The methodology of our study can be summarized as follows:

1. **Data Accumulation**: Select countries to study and collect local and international news articles specific to these countries.

2. **Data Processing**: Prepare the accumulated raw dataset for analysis through text processing.

3. **Topic Modeling**: Distinguish collected articles into distinct topics to perform topicwise comparisons.

4. **Sentiment and stance analysis**: Perform topicwise sentiment and stance detection though

   - *Hard classification*: Either positive / neutral / negative sentiment and for / impartial / against country stance.
   - *Soft classification*: Assign sentiment and stance score ranging from -1, for negative or against, to +1 for positive or in favor stance.

5. **Hypothesis testing**: Employ statistical inference tests to determine distinctions present in the news data in order to come to a conclusion abour freedom of press.

The following subsections present each of these steps in greater detail.

3.1 Data Acquisition

For this study, we decided to move forward with three countries, Canada, China, and Russia. The countries were picked based on political inclinations. For each of these countries, we collected news from local and international sources. Table 1 presents our data sources for each countries.

Source Type	Canada	China	Russia
Local	Global News and CBC	China Daily	The Moscow Times
International	Reuters and Associated Press		

Table 1: Data sources for countries under study

We used *Selenium* and *News Please* to collect our news corpus. In particular, we crawled news websites using Selenium and collected articles URLs. Afterwards, we scraped article data using *News Please*. News Please gives us a lot of data about the articles. [8] The following are properties of interest for our study:

1. **Title**: The title of the articles.
2. **Description**: A short description of the article. Typically, the subheading of the article.
3. **Maintext**: The body of the article.
4. **Publication Date**: When the article was published.

Our accumulated data and their countrywise distribution is shown in figure 1.

3.2 Data Processing

To prepare the data for topic modeling and sentiment-stance analysis we had to process the dataset. This inlcludes the following:

1. **Imputation**: Many of the URLs scraped by News Please, some attributes were null. The first step was filling in those values. We chose to replace null text columns with empty strings.

Figure 1: Collected data across countries of interest from local and international sources

2. **Duplicate removal**: In this step, we detected duplicate article data and removed them. In particular, we detected duplicates by comparies the *publication data* and *title* attributes.

3. **Text level processing**

   - Remove HTML tags, links, emails, phone numbers, etc.
   - Discard editorial information, e.g. "Published by Global News".
   - Strip non-ASCII characters.
   - Remove editorial information, e.g. written by X, photo by Y, etc.

4. **Row level processing**

   - Drop rows with empty titles and maintext.
   - Filter out unrelated geographical news.
   - Remove rows where the maintext is too short.

5. **Country level processing**

   - Global news in particular reports inflation news. This sort of articles don't have a lot of text and is strictly tabular. We decided not to include this sort of news in our analysis.
   - Removed Israel-Palestine from all country news corpus. As they are not relevant to the countries being studied.

3.3 Topic Modeling

BERTopic is an unsupervised machine learning algorithm for topic modeling. It leverages Bidirectional Encoder Representations from Transformers (BERT) and c-TF-IDF (Class-Based Term Frequency - Inverse Document Frequency) to create coherent and easily interpretable topics, described by automatically generated labels. It has a number of sub-components, which are discussed below, whose understanding at a higher level can inherently improve the topic modeling performance. [6]

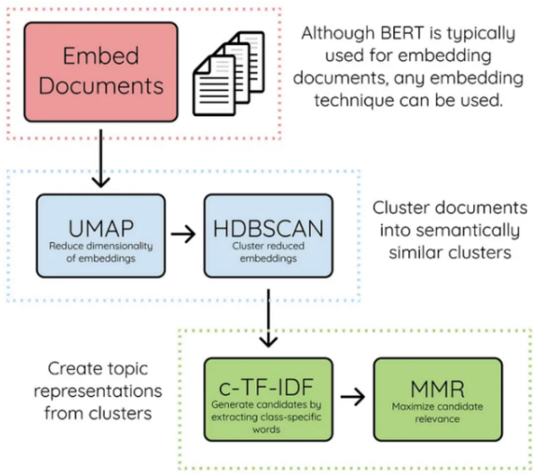


Figure 2: BERTopic Diagram

3.3.1 Transformer Embedding Model

The model embeds Input documents to sentence vector representations which are typically dense in nature. In this project, we used sentence transformers to capture semantic similarity, as it provides an extensive library of high performing sentence embeddings. The L6-v2 version created 384-dimensional sentence embeddings.

3.3.2 Dimensionality Reduction

As a next step, BERTopic then performs dimensionality reduction of the 384-dimension sentence embeddings into a lower-dimensional space, either two or three-dimensional vectors. This project used the BERTopic default; Uniform Manifold Approximation and Projection (UMAP). UMAP captures both the local and global high-dimensional space in lower dimensions.

Since BERTopic provides room for independence between steps, other popular choices for dimensionality reduction such as Principal Components Analysis (PCA) and t-distributed Stochastic Neighbour Embedding (t-SNE) can be used.

PCA works by preserving larger distances by mean squared error. This results in the data's global structure being reserved; a pro when there are clusters that are easily distinguishable in the dataset. A demerit of using PCA is that it falls short when dealing with more nuanced data where local structures are vital. Inversely, t-SNE main characteristic is the preservation of similarity. UMAP exploits the best of both worlds, thus an appropriate technique.

3.3.3 Clustering

Involves clustering dimensionally reduced embeddings into groups of similar embeddings to extract topics. There are an array of clustering techniques such as partition-based, hierarchical based and density-based, to name but a few. Each one of them has its own advantages and disadvantages.

Hierarchical Density Based Spatial Clustering of Applications with Noise (HDBSCAN) is both hierarchical and density-based. It is BERTopic's default that is typically used to capture structures with different densities. Advantages of HDBSCAN include: ease of tuning and visualizing hierarchical data, outlier identification and the ability to cluster irregular shapes.

Since BERTopic's independence between steps still hold, K-means can be used for clustering using a centroid approach. A drawback would be poor handling of noisy clusters.

3.3.4 Vectorizers

As a final step of the BERTopic workflow, topic extraction from the clusters formed in the previous step is done. BERTopic applies the use of a modified version of TF-IDF called Class-Based Term Frequency – Inverse Document Frequency (c-TF-IDF), which finds the most essential and relevant terms given to each document within a cluster. A CountVectorizer is initialized, and together with TF-IDF, topic representations are created, allowing for parameter fine tuning.

3.3.5 Topic Quality

After tuning the different parameters, we were able to generate topics for our collected data. Figure 3 shows the quality of the topic models. A good topic model will have topics that have low intertopic similarity. Moreover, the bubbles in the intertopic distance map should be large and non-overlapping. The figures clearly show this for our modeled topics. Here are a few interesting observations:

1. Canadian news has the most amount of topics. Moreover, most of the topics seem to be quite dissimilar to one another.

2. Chinese news topics have more similarity between topics. But the media covers a wide array of topics in general.

3. Russian news media seem to cover more political news. This is expected as there is a war that's ongoing.

3.4 Sentiment and Stance Analysis

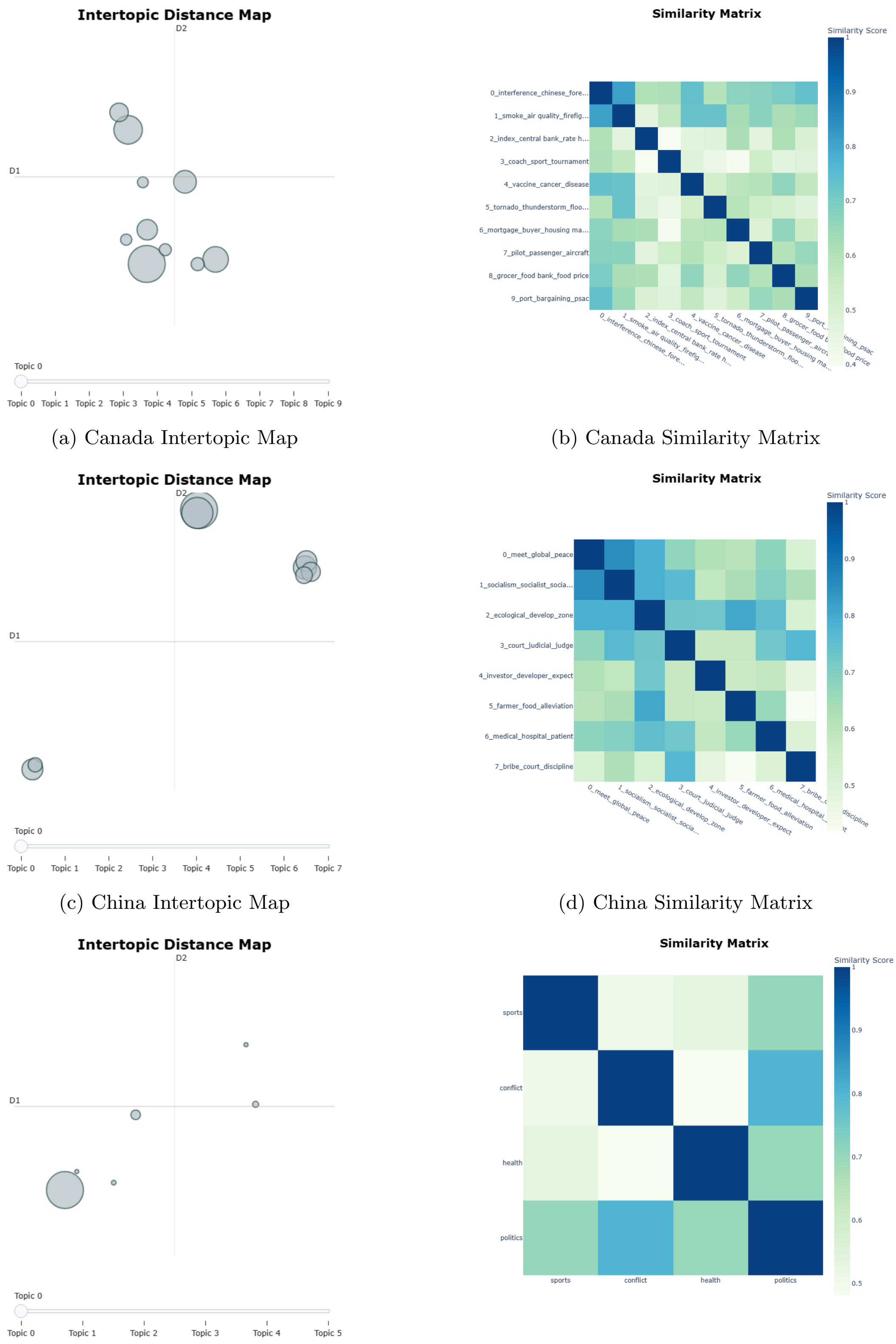
For sentiment and stance analysis, we decided to use a Large Language Model (LLM).

3.4.1 Utilization of LLM

LLMs are excellent for natural language understanding and generation tasks. [3] We used these capabilities of an LLM in the following way:

1. **Classification**: Classify the document's sentiment as positive or negative or neutral. Classify the stance as for or neutral or against the local country.

2. **Scoring**: Assign a score from -1.0 (negative / against) to +1.0 (positive / for) for sentiment and stance.



(a) Canada Intertopic Map

(b) Canada Similarity Matrix

(c) China Intertopic Map

(d) China Similarity Matrix

(e) Russia Intertopic Map

(f) Russia Similarity Matrix

Figure 3: Intertopic distance maps and similarity matrices for accessing quality of topic models

3. **Reason**: The LLM gives a short one sentence reason behind its classification and scoring.

We decided to use the open-source **LLaMa-2** model, from *Meta*. For our training and inference pipeline we heavily relied on *HuggingFace*. [14]

3.4.2 Finetune LLaMa-2

To finetune our model, we select 300 samples randomly. 100 samples for each country, and 50 samples for each source type. Afterwards, we engineered an effective prompt using *Prompt Perfect* for LLaMa-2. The prompt is shown in listing 1. It should be noted that we replaced {*title*} and {*content*} with the actual title and content of the news article being analyzed.

```
As a neutral news analyst, assess the sentiment and stance of the news article
excerpt and assign a score between -1.0 (completely negative/against-{country})
and 1.0 (completely positive/pro-{country}) for both sentiment and stance.
Provide a single short sentence to justify your scores, drawing on the article's
language, tone, and presentation to support your analysis.

Article Excerpt:
- Title: "{title}"
- Content: "{content}"

Output format:
1. Sentiment: [Positive/Neutral/Negative]
    * Score: [Your Score]
    * Reason: [Your Reason]
2. Stance: [Pro-{country}/Impartial/Against-{country}]
    * Score: [Your Score]
    * Reason: [Your Reason]

```

Listing 1: Prompt for LLaMa

To create the finetuning dataset, we fed the 300 articles into ChatGPT using the prompt structure shown above. The responses from ChatGPT were manually varified for validity. Afterwards, the responses were accumulated to create the finetuning dataset. The finetuning dataset needed to be fed into the LLaMa-2 model in a particular format, shown in lisitng 2.

```
###Human:
As a neutral news analyst, assess the sentiment and stance of the news article
excerpt and assign a score between -1.0 (completely negative/against-{country})
and 1.0 (completely positive/pro-{country}) for both sentiment and stance.
Provide a single short sentence to justify your scores, drawing on the article's
language, tone, and presentation to support your analysis.

Article Excerpt:
- Title: "{title}"
- Content: "{content}"

Output format:
1. Sentiment: [Positive/Neutral/Negative]
    * Score: [Your Score]
    * Reason: [Your Reason]
2. Stance: [Pro-{country}/Impartial/Against-{country}]
    * Score: [Your Score]
    * Reason: [Your Reason]

###Assistance:
```

Listing 2: Input for LLaMa

Next, we finetuned the base model from HuggingFace user *TinyPixel* in the following way:

- **Sharded Base model**: A sharded base model allowed us to load chunks of the model one at a time. This, along with QLoRA, allowed us to train the model on a single GPU in google colab.
- **QLoRA**: We used QLoRA to quantize the model and train a 4-bit version of the LLM. Simply, this allows us to train a subset of the model's parameters instead of the entire model. This allowed us to finetune the model with only a single GPU. [4]
- **PEFT**: Parameter Efficient Fine Tuning was used to train a subset of the model's parameters instead of the entire model. This allowed us to train the model quickly while retaining high performance.

After finetuning the model, 4000 samples were selected for each country. Then the sentiment and stance of those countries were predicted using LLaMa-2.

3.5 Hypothesis Testing

After getting the scores for our dataset, we performed statistical analysis on the scores of the news. In particular, we performed the following hypothesis tests:

Test	**Parameter of Interest**	**The Question We Ask**
Welch Test	Mean	Does the average score from both sources have any difference?
Wilcoxon Test	Median	Is the meidan score from sources different?
Variance Test	Varaince	Does the content from both sources cover a similar range of sentiment or stance?
Pearson's Test	Correlation	Is there any linear correlation between the scores from both sources?
Spearman's Test	Monotonic Relation	Is there any monotonic relationship between the scores from both sources?

Table 2: Role and Goal prompt engineering and output

All tests were done with a confidence level of 99%. Since we have a lot of data, we wanted to be sure to come to a conclusion.

4 Experiment

We break our expriment into 3 case studies. Each case study covers a specific country and analyzes the news from both local and international sources.

4.1 Canada

From figure 4, we can see than Canadian news covers a wide range of topics.

4.1.1 Sentiment Analysis

We first begin by showing the distribution of the sentiment scores for each topic from both the local and international media in figure 5.

Some distinctions are clearly visible here. For example:

- *Sports* news from both sources tends to have a positive sentiment. Both sources seem to have the same sentiment in this topic.
- *Telecommunication* news from international media seem to cover a wide spectrum of sentiment while local media seem to convey a more negative portrayal of the matter. *Public-service* has the same distribution while *Crime* has the opposite.
- *Transportation* and *weather* has a noticable distinction in their score distribution for both sources.

We run statistical inference tests to find these distinctions mathematically. The results are shown in figure 6.

- *Crime*, *transportation*, and *weather* is statistically significant in terms of Welch and Wilcoxon test. Which means that on average the sentiment scores from both sources are different. This means both sources typically report news that are of different sentiment.
- *National-security* is statistically significant in terms of Pearson's and Spearman's test. Which means the score of the two sources aren't independent of each other. There exists some sort of a correlation or monotonic relationship. This means that the two sources report news that are somewhat related.

4.1.2 Stance Analysis

Now let's analyze the stance scores from figure 7. The most interesting observation here is that, international media, in most topics, seem to have an impartial stance, While local news outlet has a more pro-canada stance. The obvious exceptions are *Health* and *Food*.

From the statistical test results shown in 8, we see a similar result as the sentiment score inference results.

- *Crime* and *Sport* relatd news seem to have different stance scores on average.
- While *Conflict* related news isn't independent according to Pearson's and Spearman's test.
- Interestingly, *Finance* news has a different variance for stance scores. This means, the two sources cover a different range of stances.

4.1.3 Text Analysis

Let's look at a wordcloud for *Crime* related news from both sources in figure 9.

We can see that local and international media use the same sort of words across their negative and positive sentiment news. For example, the word police, fire, government appears in the negative sentiment word cloud for both sources. On the other hand, indigenous community, first, fire, etc. are common in the positive sentiment.

Figure 4: Canadian news topics bar chart.

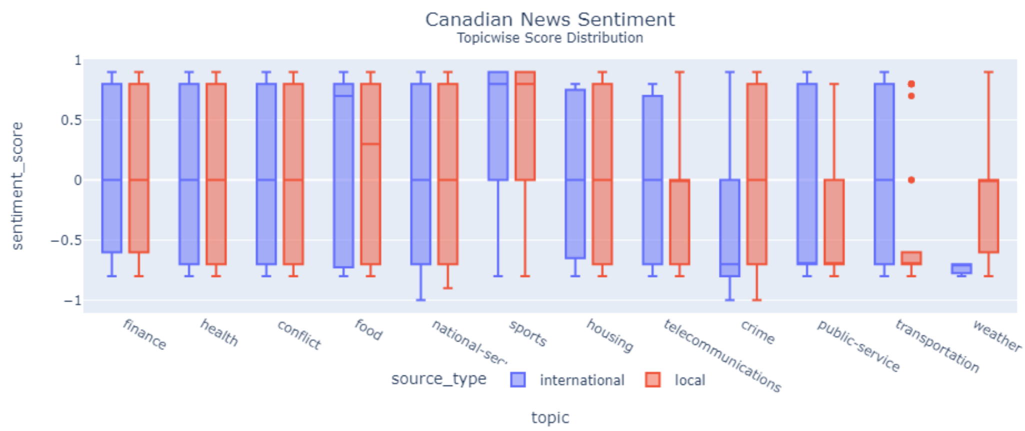


Figure 5: Boxplot showing the topicwise distribution of canadian news sentiment scores for local and international media outlets.

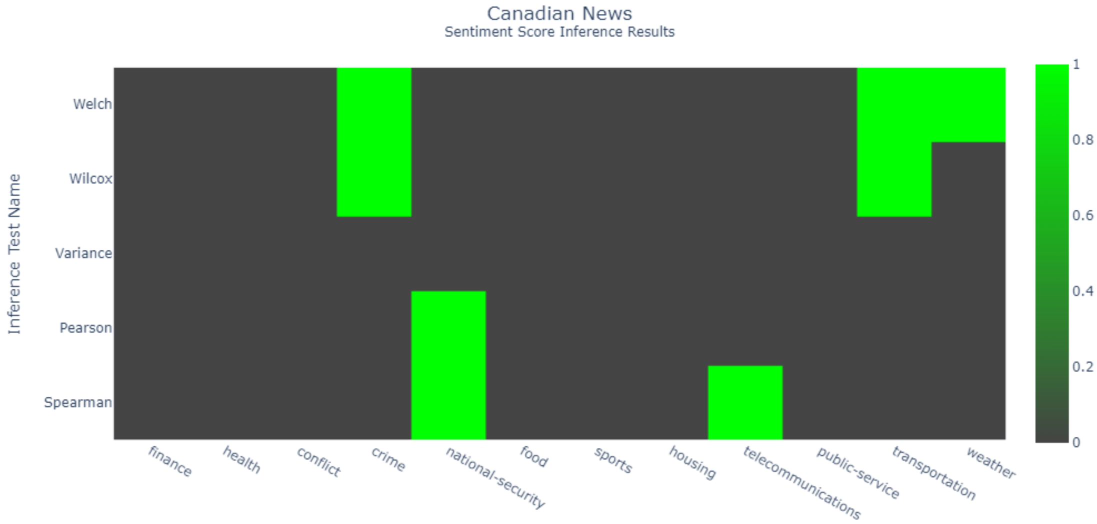


Figure 6: Heatmap showing the result of topicwise inference tests done on canadian news sentiment scores.

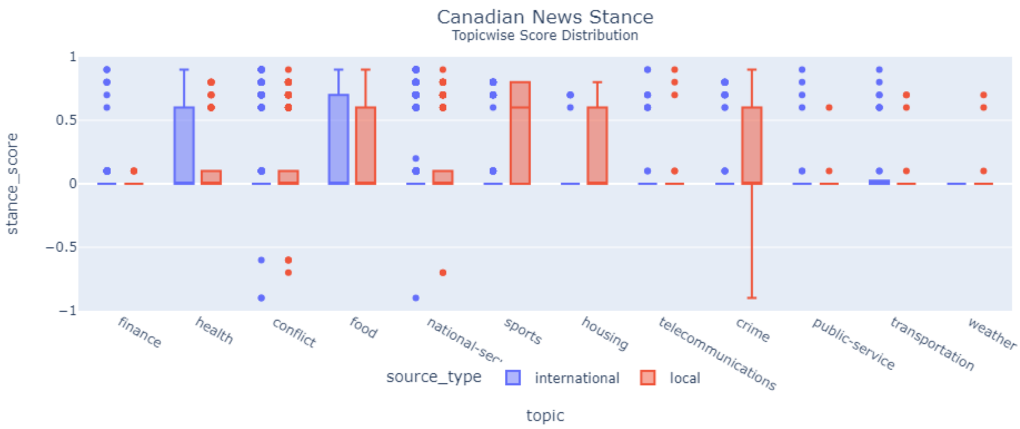


Figure 7: Boxplot showing the topicwise distribution of canadian news stance scores for local and international media outlets.

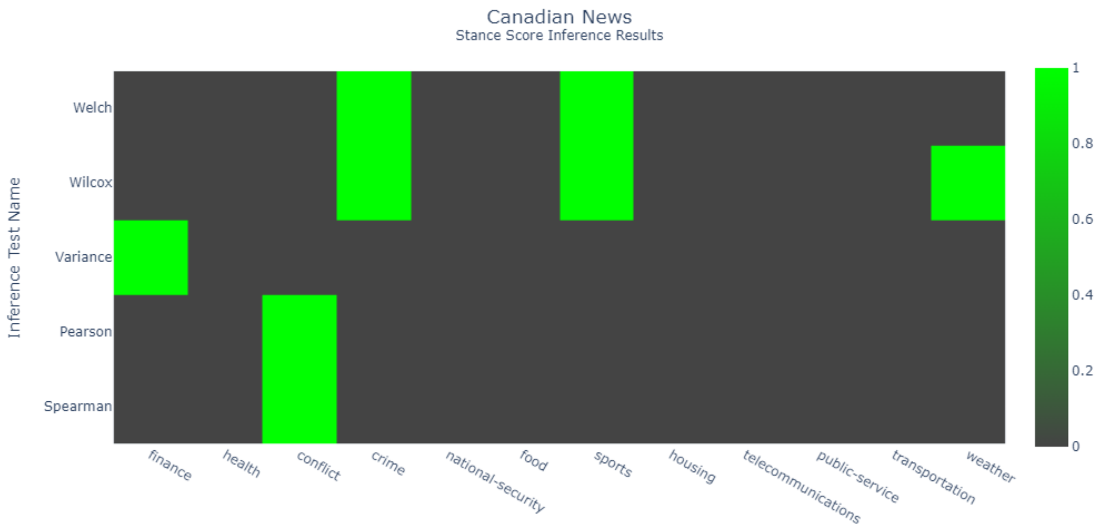


Figure 8: Heatmap showing the result of topicwise inference tests done on canadian news stance scores.

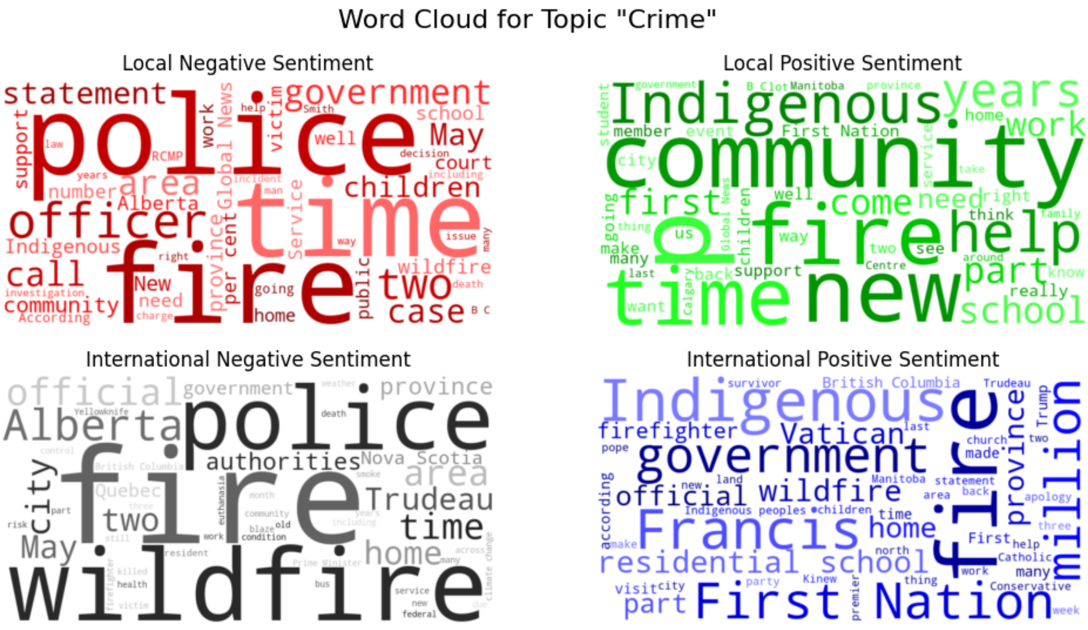


Figure 9: Word cloud for canadian news about crime.

4.1.4 Verdict

Although there is statistically significant distinctions in sentiment for a few topics, both sources report news that are similar in sentiment and stance. Which is why, we find canadian news organization "Global News" as a free press organization.

4.2 China

Let us now look at Chinese news media outlet. Figure 10 shows the distribution of news amongst different topics from both sources. It's clear that both sources publish more news about *politics* than any other topic. The international media also publishes more news about *finance* and *international-relations* than the local news media as well. However, *agriculture* related news is covered mostly by local news outlet.

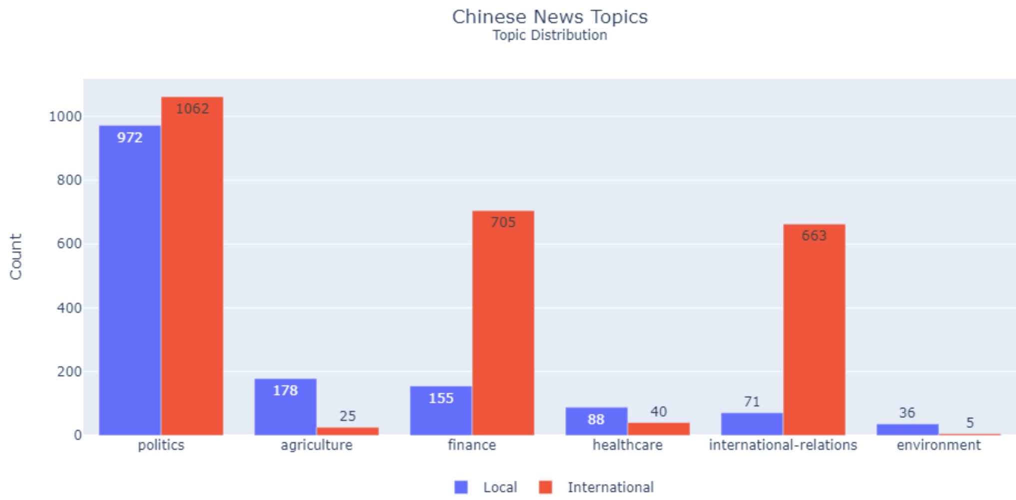


Figure 10: Chinese news topics bar chart.

4.2.1 Sentiment Analysis

Figure 11 shows the distribution of sentiment scores for various topics for local and international media.

One of the most interesting observation is that, the local media in most cases always seems to cover positive sentiment news. Except for a few outliers, their distribution is concentrated towards positive scores. On the other hand, international news covers a wide spectrum of sentiment.

Figure 12 confirms the distinction we saw in the sentiment score distribution. Not only does most topic have on average a different sentiment score, they also have a very different score variance as well. Safe to say that the local and international media cover very different set of news when it comes to sentiment.

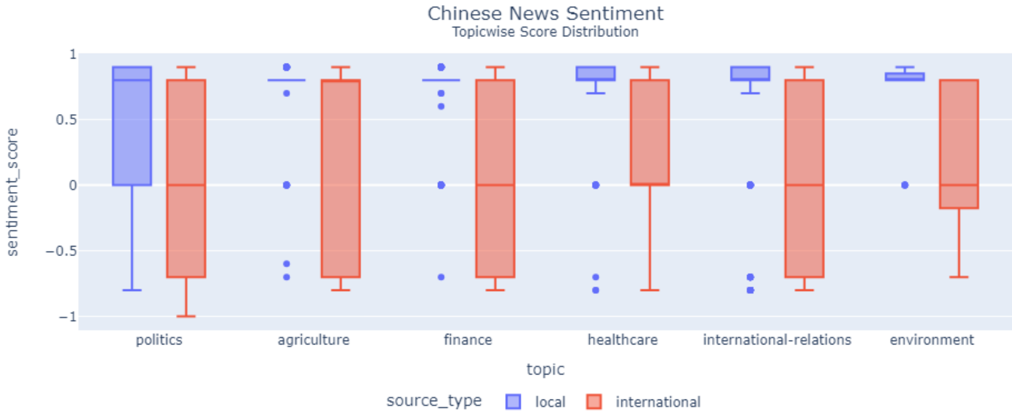


Figure 11: Boxplot showing the topicwise distribution of chinese news sentiment scores for local and international media outlets.

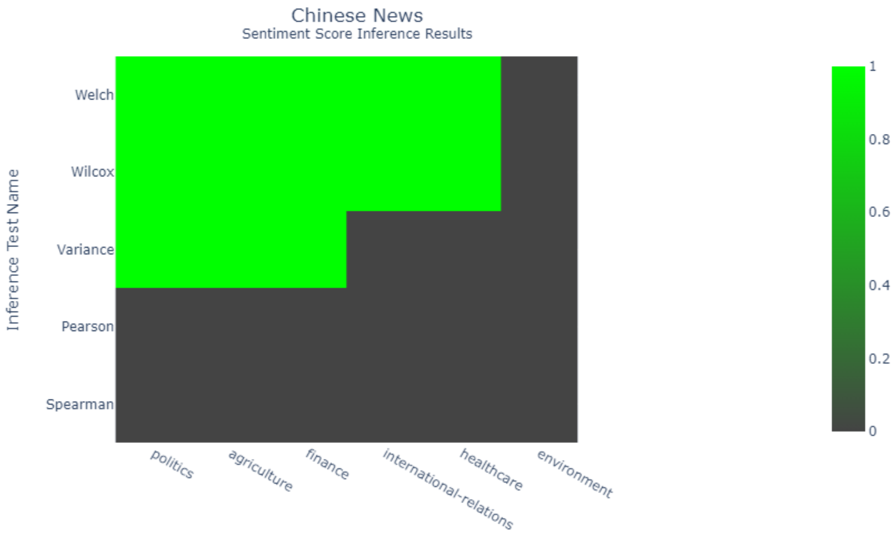


Figure 12: Heatmap showing the result of topicwise inference tests done on chinese news sentiment scores.

4.2.2 Stance Analysis

From figure 13, we observe a similar trend. Local news seem to trend towards pro-china stance while international media is mostly impartial, except for a few outliers. However, local news media seem to cover a few against china news when it comes to *politics* and *international-relations*. On the other hand, *environment* is covered in an impartial to slightly more against-china stance by the international news media.

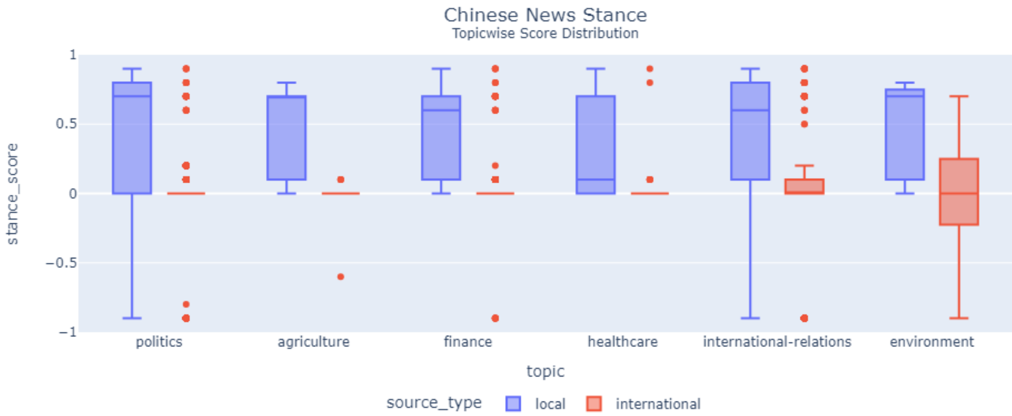


Figure 13: Boxplot showing the topicwise distribution of chinese news stance scores for local and international media outlets.

Our observations are confirmed when we see the hypothesis tests results in figure 14. Most of the topics, except for *enviornment*, have a statistically significant difference in stance scores. The only reason we don't have a statistically significant difference is that we don't have enough data for this topic. We can also see than most topics are also independent as they are not significant in terms of pearson's and spearman's correlation.

4.2.3 Text Analysis

The word cloud for *politics* is shown in figure 15. Unlike the wordcloud for canadian news, there are less things in common here. For example, in international news, *Beijing*, *government*, and *official* is highlighted in the negative news. While for local news, *US*, *law*, and *court* are highlighted more in the negative sentiment news. International media also talks a lot about *Beijing* and *government* in positive light as well. But the local media seems to talk about *CPC* and *Central Committee* more. Interesting to see that the local media barely has any mention of *Beijing* when the international media is talking about it so much.

4.2.4 Verdict

There seem to be a lot of dissimilarity when it comes to local news media outlet, "China Daily", and the international news media outlets. Statistically significant differences can be found across most topics. Which is why, we fail to say "China Daily" has freedom of press based on the news samples we have collected.

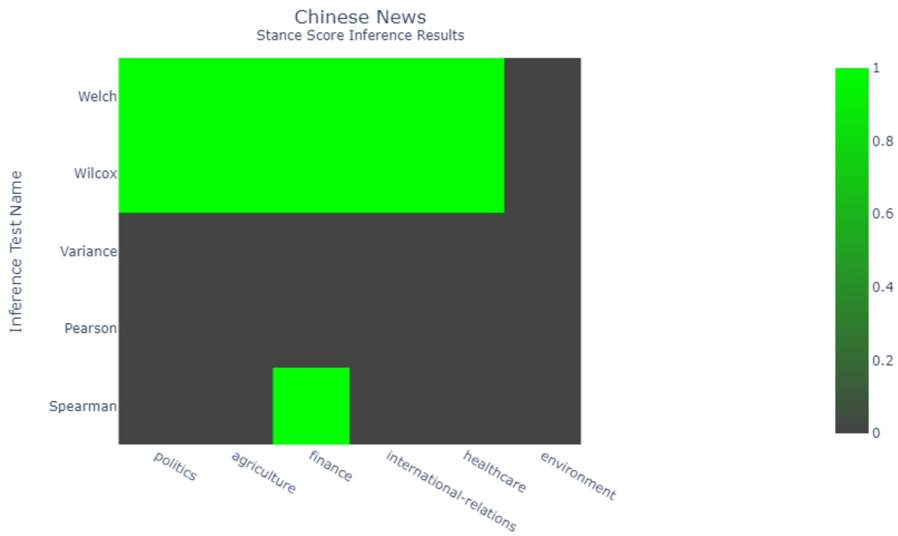


Figure 14: Heatmap showing the result of topicwise inference tests done on chinese news stance scores.

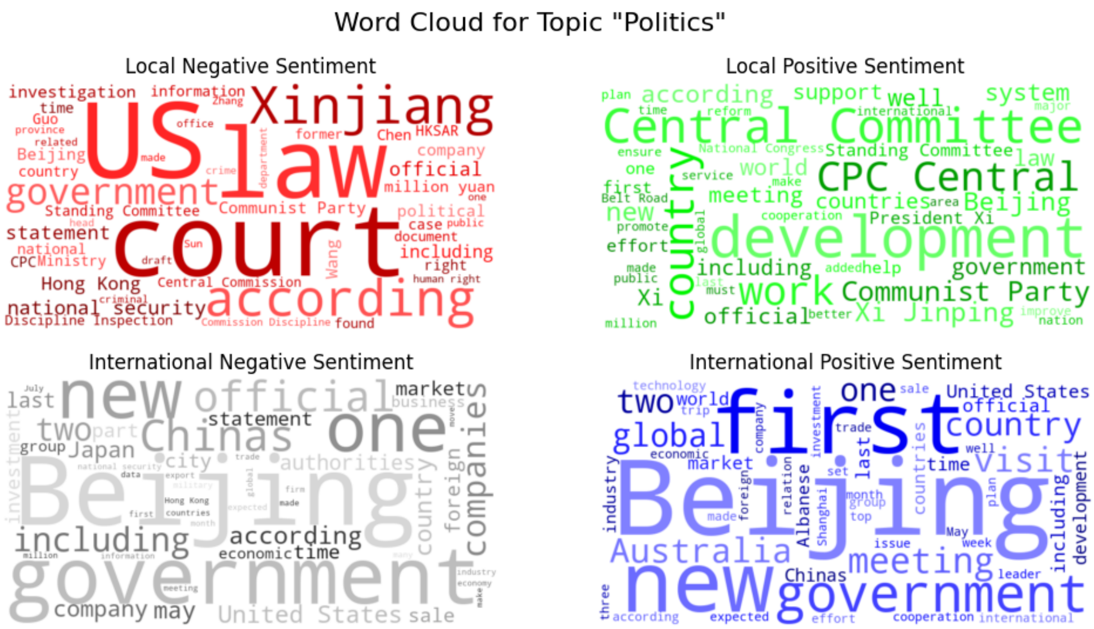


Figure 15: Word cloud for chinese news about politics.

4.3 Russia

Finally, we analyze russian news. Figure 16 shows the distribution of news covered by both local and international media outlets. An interesting observation is that both sources publish more news about *politics* than all other topics combined. Moreover, local news covers more about *health* related news, while international news dives more into the ongoing *conflict* news.

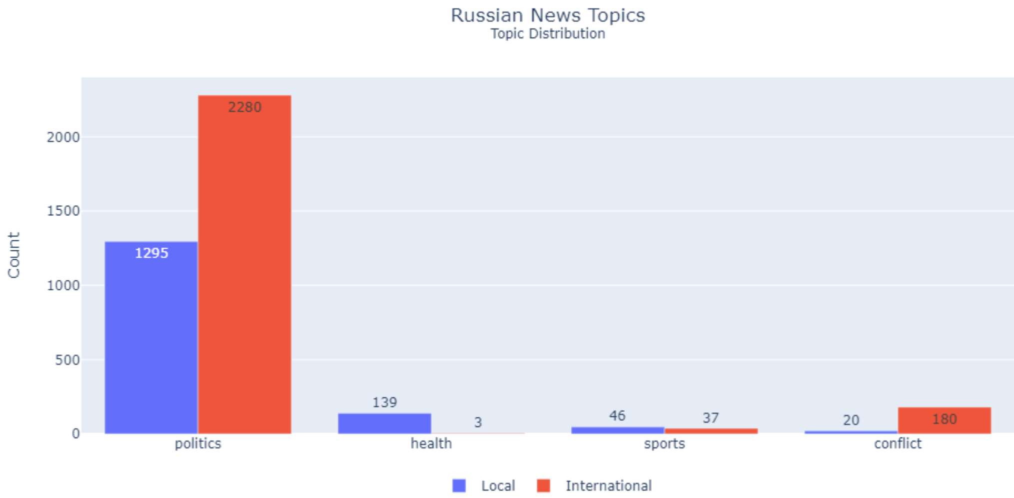


Figure 16: Russian news topics bar chart.

4.3.1 Sentiment Analysis

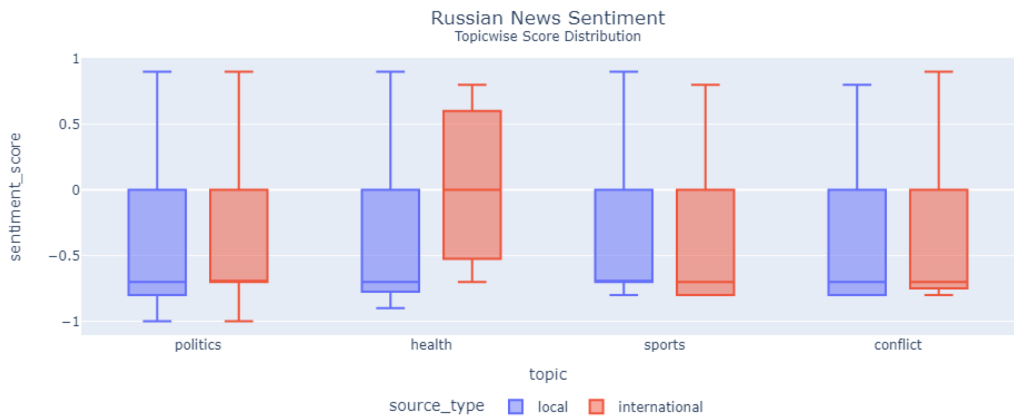


Figure 17: Boxplot showing the topicwise distribution of russian news sentiment scores for local and international media outlets.

From figure 17 we can see that, unlike chinese news, there are less distinctions in the distribution of the news articles from both sources. *Health* related news seems to be more negative in local media, while the international media is more neutrally distributed. Both new media organizations seem to convey a similar negative sentiment when it comes to *conflict* and *politics*.

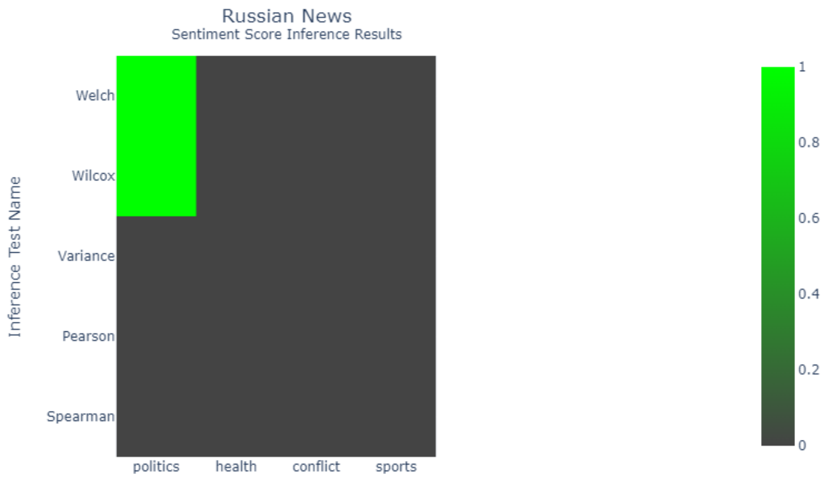


Figure 18: Heatmap showing the result of topicwise inference tests done on russian news sentiment scores.

A more interesting observation comes from the statistical test as graphed in figure 18. There seems to be a statistically significant difference on the average and median sentiment score between the two sources.

4.3.2 Stance Analysis

The stance scores for both sources seem to be mostly impartial. Except for *Health* related news, where the international media seems to be trendly slightly towards a pro-russian stance.

The inference tests for stance, tells the same story as that of the sentiment. Political news on average has different stances from both sources. Interestingly, *conflict* related news has a statistically significant difference in variance. Meaning, local and international news doesn't cover the same range of stance when it comes to this topic. Also, *sports* news has a monotonic relationship between the two sources.

4.3.3 Text Analysis

Analyzing the word clouds shown in figure 21, we can see that *Ukrain* is a huge talking point across all sentiments regardless of the source. *Putin* also comes up more in local news while *United States* and *China* comes up a lot in international news. In particular, the US seems to be associated more with negative sentiment while China is associated with more positive sentiment. This could indicate the alliance Russia has with China and the animosity it has with the US. Overall, the distinctions in textual content isn't quite as clear as with chinese news.

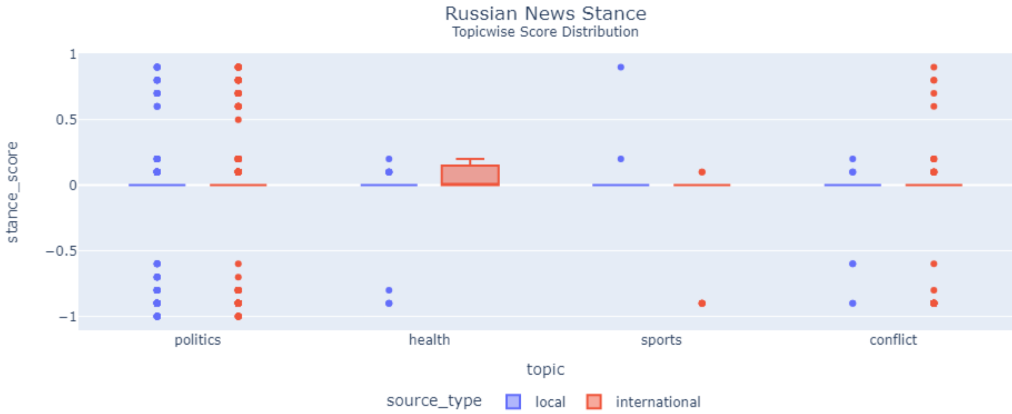


Figure 19: Boxplot showing the topicwise distribution of russian news stance scores for local and international media outlets.

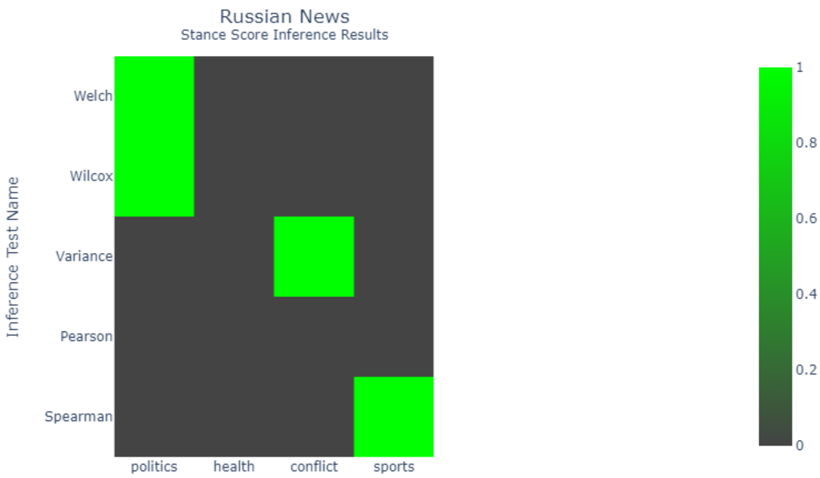


Figure 20: Heatmap showing the result of topicwise inference tests done on russian news stance scores.

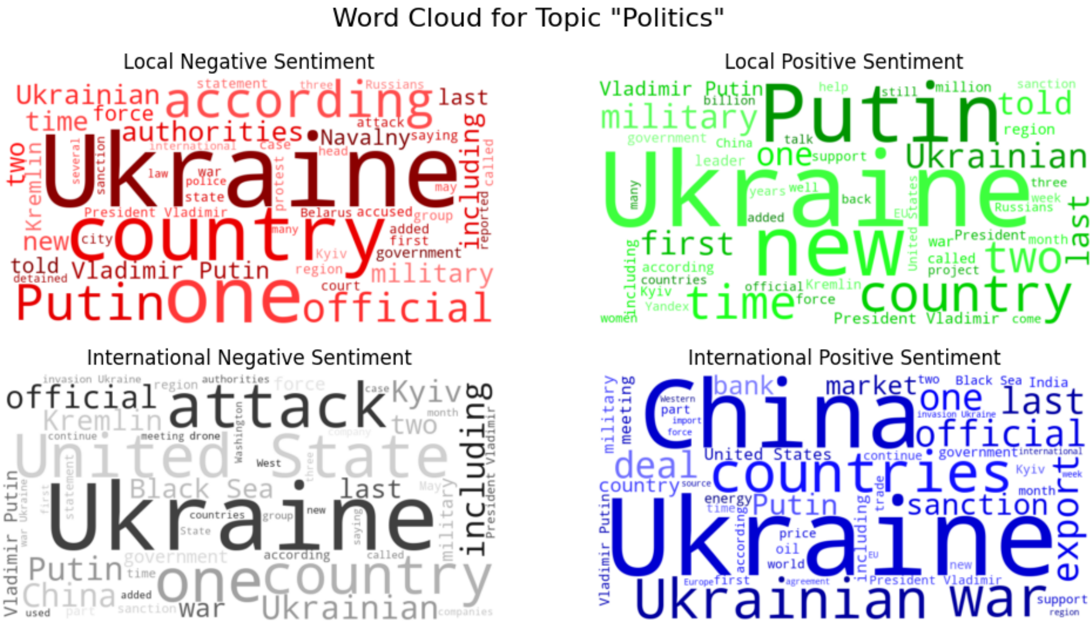


Figure 21: Word cloud for russian news about politics.

4.3.4 Verdict

Overall, we fail to find distinctions across all topics. This is mainly due to the lack of sufficient data for topics other than *politics*. As a result, we are unsure, as to whether "The Moscow Times" has freedom of press or not. From the sample data we have, it seems like the local outlet mostly agrees with the international outlets. However, due to limited samples for other topics, we can't say this with certainty.

5 Limitations

1. The biggest limitation is the lack of a sufficiently large dataset. For *Russia*, we saw that for most topics, we simply don't have enough data.
2. Another huge limitation comes form the LLaMa-2 model.
   (a) LLaMa-2 was finetuned on a very small dataset of 300 samples.
   (b) The model seems to output scores that trend towards either extremes and very rarely towards the the middle region of positive or negative.
   (c) Due to the limitation of prompt length accepted by the LLaMa-2 model, we only gave it the first 1024 characters of an article. This affects the sentiment and stance of the article. Our assumption is that, an article's sentiment and stance is clear from the first few sentences. This assumption wasn't tested to ensure validaity on our end.
3. Not all negative news is country specific. For example, news about the two ongoing conflicts, enviornmental distress, etc. are negative news regardless of the countries stance on the matter. This wasn't considered when doing the hypothesis testing.

4. Correlation and monotonic relationship wasn't tested for directionality. Meaning, we know whether the scores are independent or not. If they are not independent, we don't know how exactly they relate. Additional testing is needed to ascertain whether the relationship is positive or negatively correlated.

6 Conclusion

Despite the drawbacks, the project is able to find distinctions present amongst news articles published by national and international news media outlets. From the case studies, we can see that these distinctions clearly show up in the hypothesis tests. However, more work is needed to make the overall process more rigorous.

References

[1] Auksė Balčytienė. Crisis of agency in central and eastern europe: From the consolidation of media freedom to the institutionalisation of free choice. *Javnost-The Public*, 28(1):75–89, 2021.

[2] Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. International sentiment analysis for news and blogs. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 2, pages 19–26, 2008.

[3] Daniil A. Boiko, Robert MacKnight, and Gabe Gomes. Emergent autonomous scientific research capabilities of large language models, 2023.

[4] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms, 2023.

[5] Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. Large-scale sentiment analysis for news and blogs. *Icwsm*, 7(21):219–222, 2007.

[6] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure, 2022.

[7] Pentti Haddington et al. Stance taking in news interviews. *SKY Journal of Linguistics*, 17(1):101–142, 2004.

[8] Felix Hamborg, Norman Meuschke, Corinna Breitinger, and Bela Gipp. news-please: A generic news crawler and extractor. In *Proceedings of the 15th International Symposium of Information Science*, pages 218–223, March 2017.

[9] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M Meyer, and Iryna Gurevych. A retrospective analysis of the fake news challenge stance detection task. *arXiv preprint arXiv:1806.05180*, 2018.

[10] Michael Meyen, Anke Fiedler, and Kerem Schamberger. 'it is a crime to be abusive towards the president': A case study on media freedom and journalists' autonomy in museveni's uganda. *African Journalism Studies*, 37(3):1–18, 2016.

[11] Jan Oster. *Media freedom as a fundamental right*, volume 30. Cambridge University Press, 2015.

[12] Krzysztof Rybiński. Political sentiment analysis of press freedom. *Studia Medioznawcze*, 19(3), 2018.

[13] Krzysztof Rybinski et al. A machine learning framework for automated analysis of central bank communication and media discourse: the case of narodowy bank polski. *Bank i Kredyt*, 1(1):20, 2019.

[14] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models, 2023.