



# MEDIA REPRESENTATION OF THE PALESTINE-ISRAEL CONFLICT

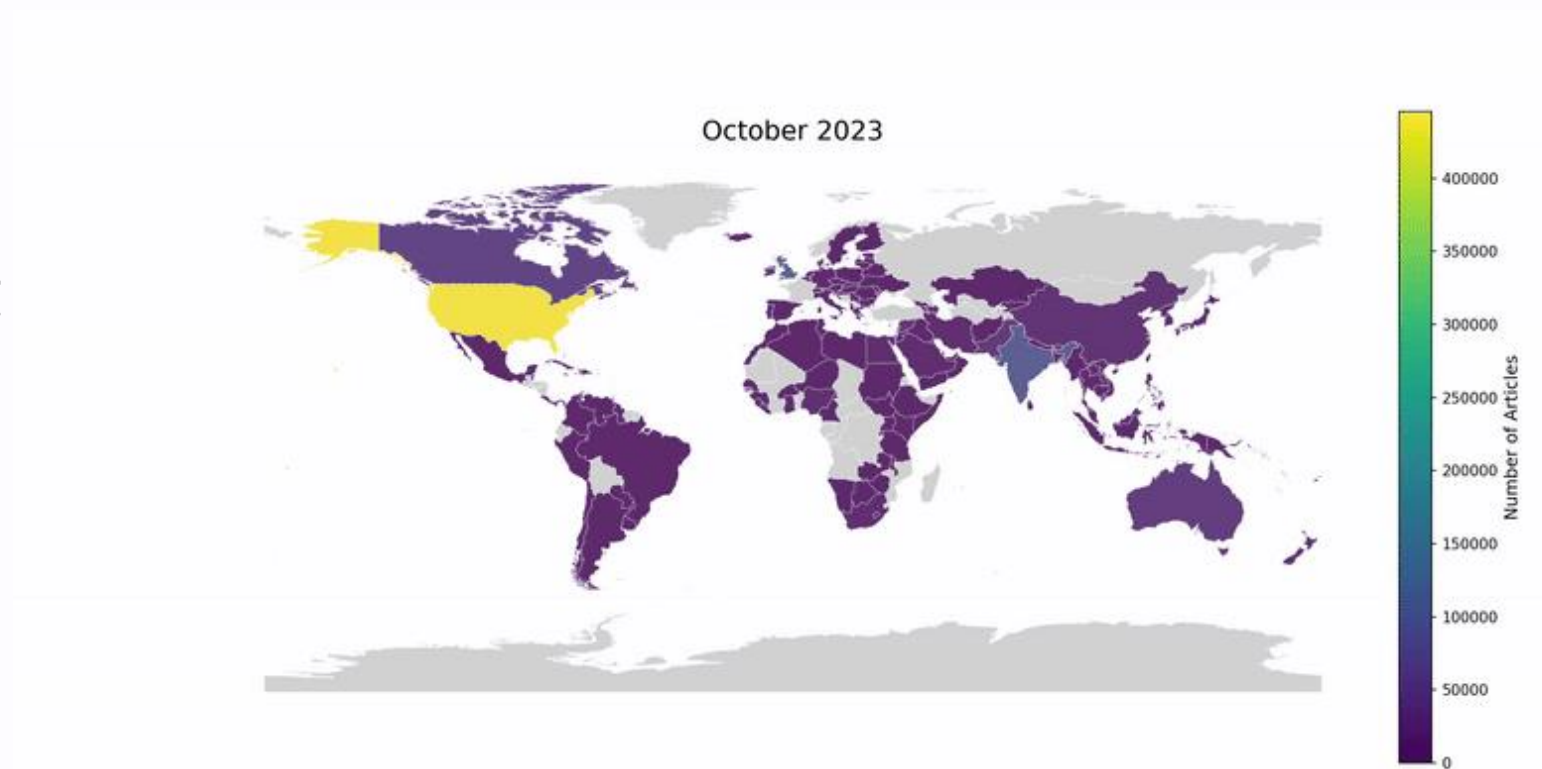
A Study of GDELT Database

David Valdivieso Lopez, Abu Taher, Kaisa Anttila, Rabiul Islam

# Background

---

- How does the media reporting of major world events vary?
- Information bubbles
- Polarizing topic: Palestine-Israel conflict



*Figure 1: Changes of news volume around the globe*

# GDELT



- Global Database of Events, Language, and Tone
- News from around the globe
- Many attributes available per news item:
  - Date and location
  - Actors
  - Tone
  - Goldstein Scale
- GDELT 1.0 used

CAMEO	192
Name	Occupy territory
Description	Occupy, seize control of a territory using armed forces.
Example	Vietnamese-led forces have retaken a strategic village in Western Kampuchea after fierce fighting with guerrillas who overran it late last month, Thai military sources said today.
Example	Burmese troops have captured part of a Karen guerrilla stronghold on the Thai border after five days of fighting, Thai police here said on Monday.
Example	Israel today mounted its long-threatened invasion of South Lebanon, ploughing through the United Nations lines on the coast of south of Tyre and thrusting forward in at least to inland areas.

CAMEO	193
Name	Fight with small arms and light weapons
Description	Attack using small arms and light weapons such as rifles, machine-guns, and mortar shells.
Usage Notes	Small arms include revolvers and self-loading pistols, rifles and carbines, sub-machine guns, assault rifles and light machine-guns. Light weapons include heavy-machine guns, hand-held under-barrel and mounted grenade launchers, portable anti-aircraft guns, portable anti-tank guns, recoilless ri-

# GDELT: documentation

- Extensive documentation available
- Detailed explanations of for example Cameo codes

# Our Process

- End-to-end analytical pipeline leveraging GDELT-1.0 event data
- Preprocessing steps involve filtering, transformation, categorization, merging and cleaning.
- Applied region-aware temporal, geographical, sentiment, and polarization analysis
- Integrated Python-based tools: PySpark, Pandas, GeoPandas, Statsmodels, Seaborn.

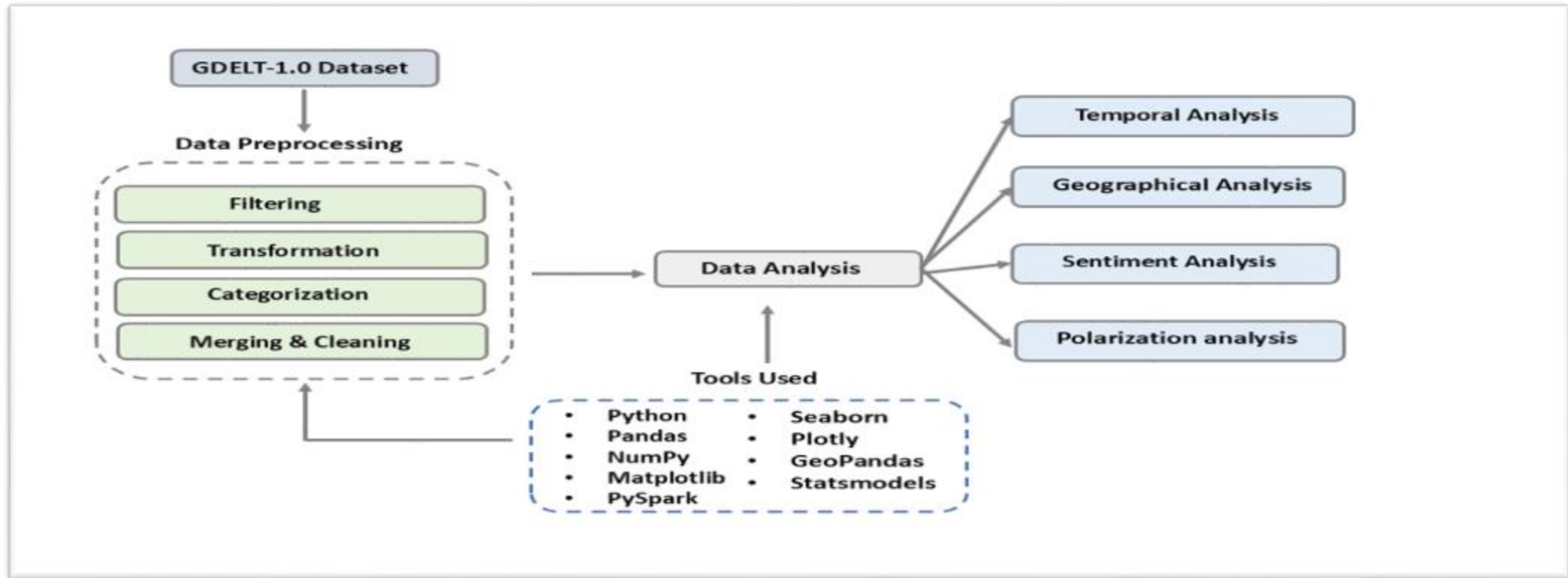


Figure 2: Overall Workflow of GDELT-Based Media Analysis Pipeline

# Cleaning the Data

- GDELT data includes events from all around the world, from several decades
- Fetching data from 6 months between October 2023 to March 2024
- Filtering the data to include keywords: Israel, Palestine, Gaza etc.
- Making data more readable: CAMEO code mapping, country codes etc.
- Data was saved as parquets to reduce 70% of used space
- Final dataset 3,5 million rows and 24 columns, about ¼ of the original

```
root
|-- Cameo: string (nullable = true)
|-- Cameo_full: string (nullable = true)
|-- GLOBALEVENTID: string (nullable = true)
|-- SQLDATE: date (nullable = true)
|-- Actor1Name: string (nullable = true)
|-- Actor1Geo_FullName: string (nullable = true)
|-- Actor2Name: string (nullable = true)
|-- Actor2Geo_FullName: string (nullable = true)
|-- IsRootEvent: string (nullable = true)
|-- EventCode: string (nullable = true)
|-- EventBaseCode: string (nullable = true)
|-- QuadClass: string (nullable = true)
|-- GoldsteinScale: double (nullable = true)
|-- NumMentions: integer (nullable = true)
|-- NumSources: integer (nullable = true)
|-- NumArticles: integer (nullable = true)
|-- AvgTone: double (nullable = true)
|-- ActionGeo_Type: string (nullable = true)
|-- ActionGeo_FullName: string (nullable = true)
|-- ActionGeo_CountryCode: string (nullable = true)
|-- SOURCEURL: string (nullable = true)
|-- domain_name: string (nullable = true)
|-- news_source_country: string (nullable = true)
|-- region: string (nullable = true)
```

*Figure 3: Schema of the final dataset*



# Analyzing the Data

- Describing data
- Polarization
- Tone
- Challenges:
  - Directionality of data
  - Transparency of certain aspects of GDELT
  - Data verification



# Results

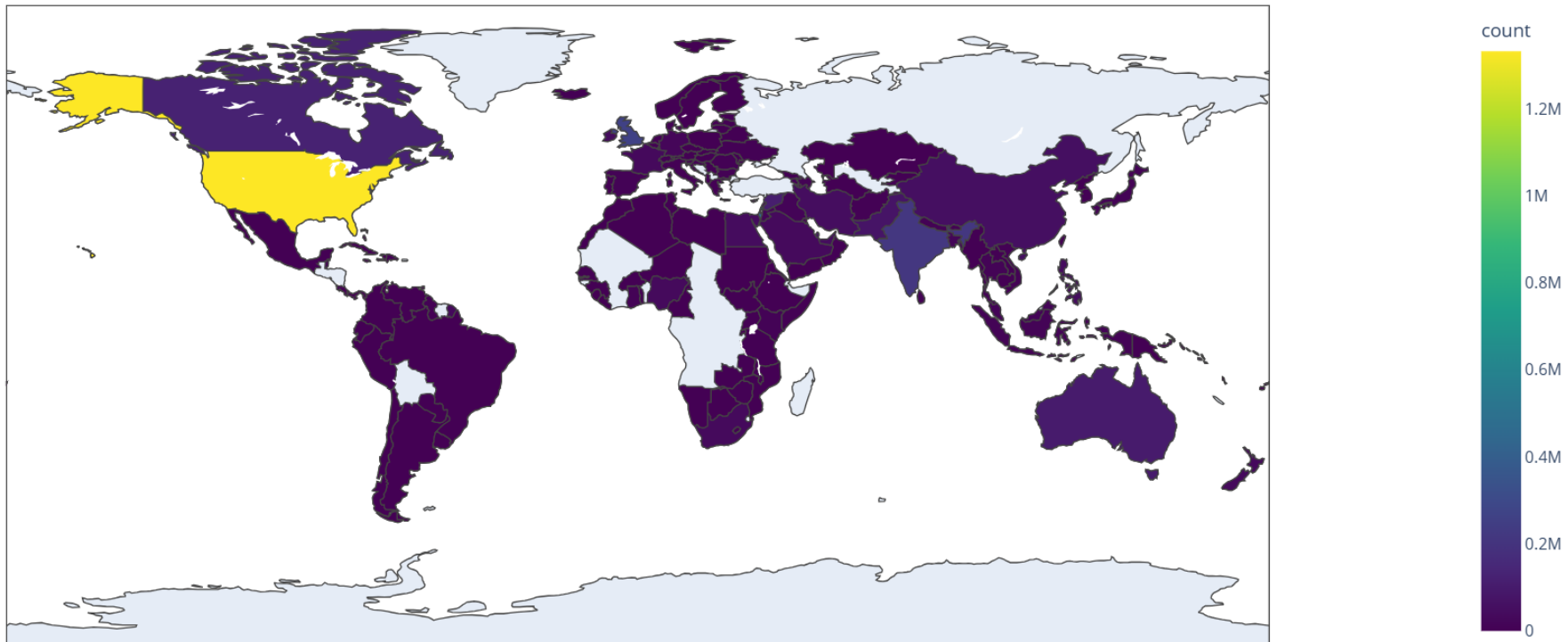
- We saw some polarization in different regions
- Some differences in tone by region and actors
- Results shown in graphs



# Describing the Data: Reporting Countries

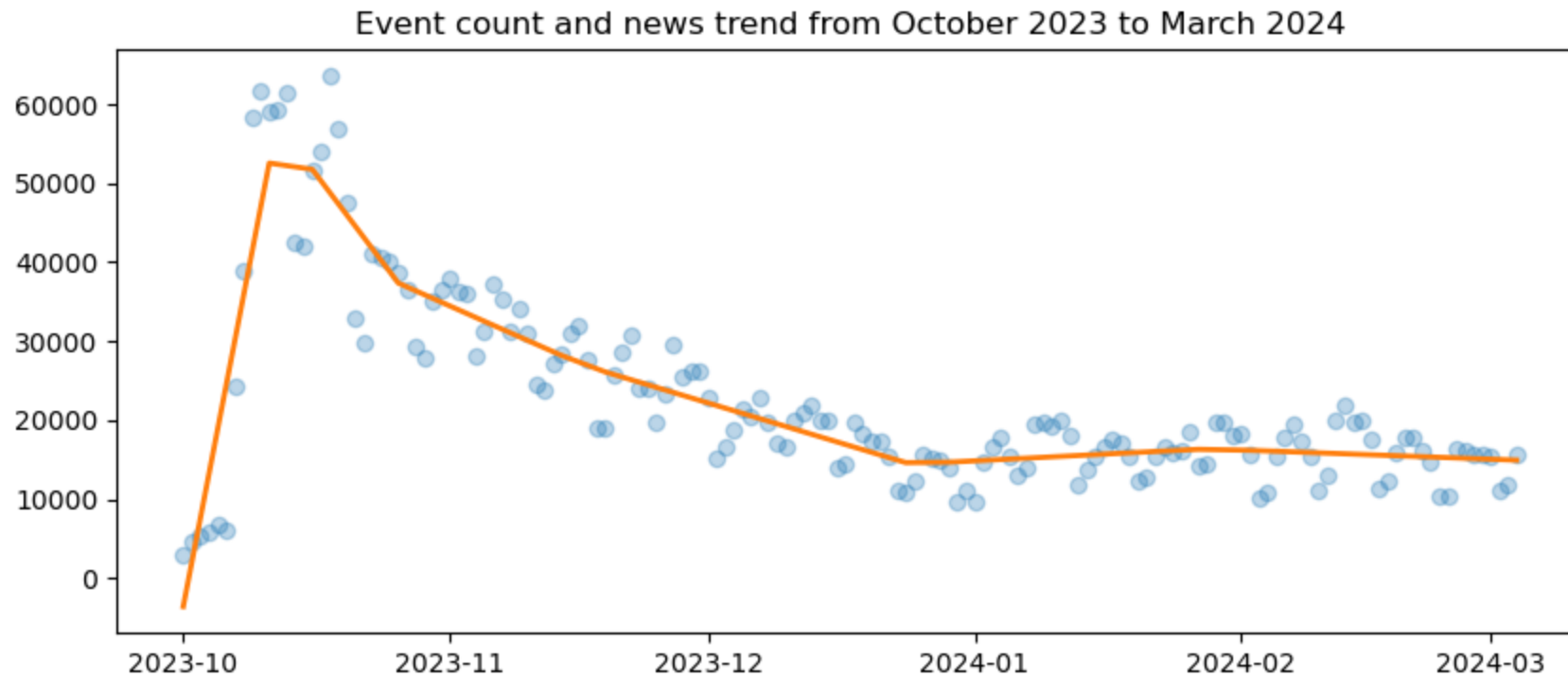
- Most articles from English speaking countries
- To solve data unbalancing, we have analyzed and compared data in different regions

News Articles by Source Country



*Figure 3: Total amount of news from all over the world*

# GDELT event time distribution



*Figure 4: Reporting trend and key event from Oct 2023 to 2024*

# Media that has reported most of the events

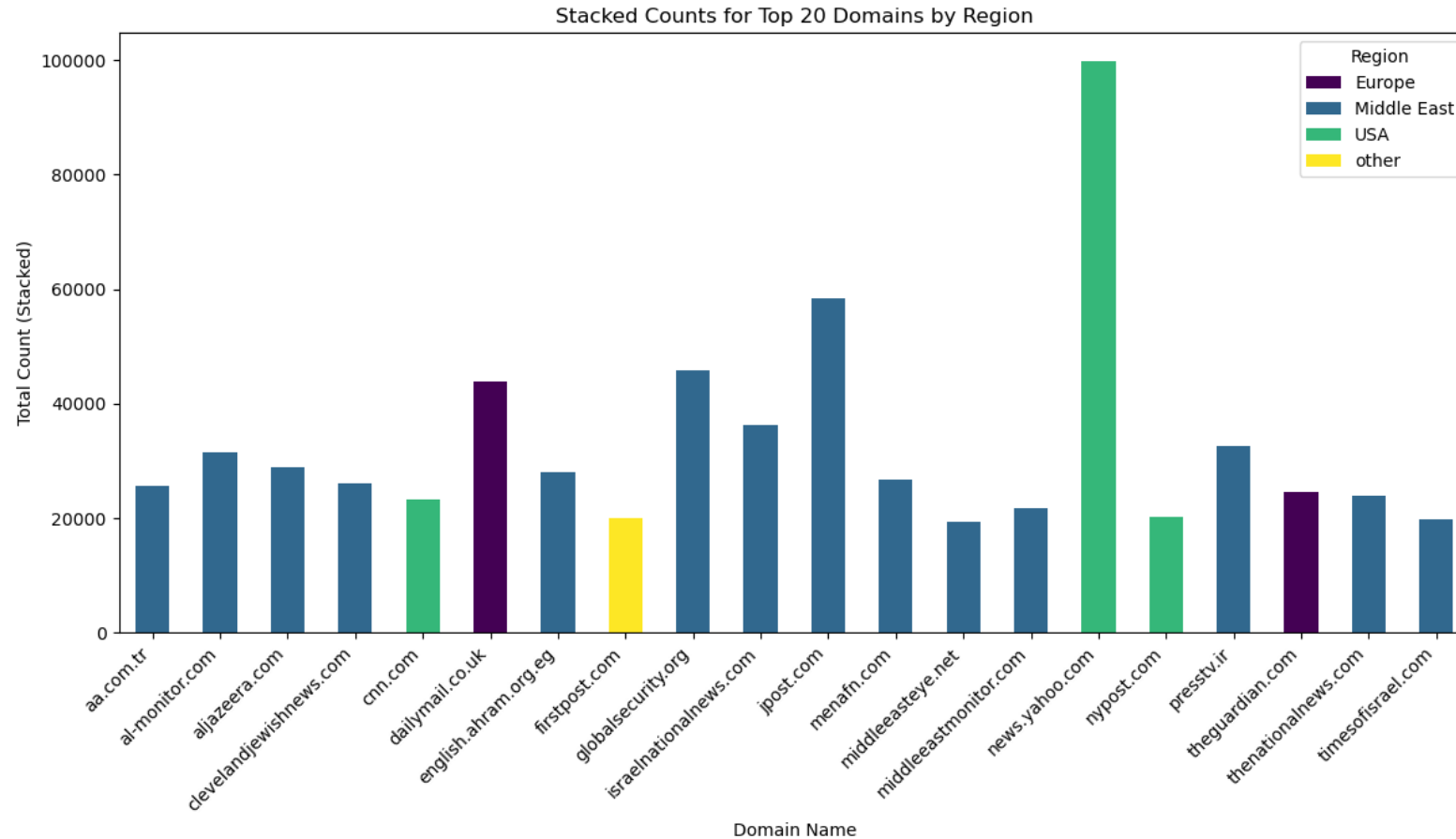


Figure 5: Stacked counts for top 20 domains by region



# Cameos reported by media in this time range

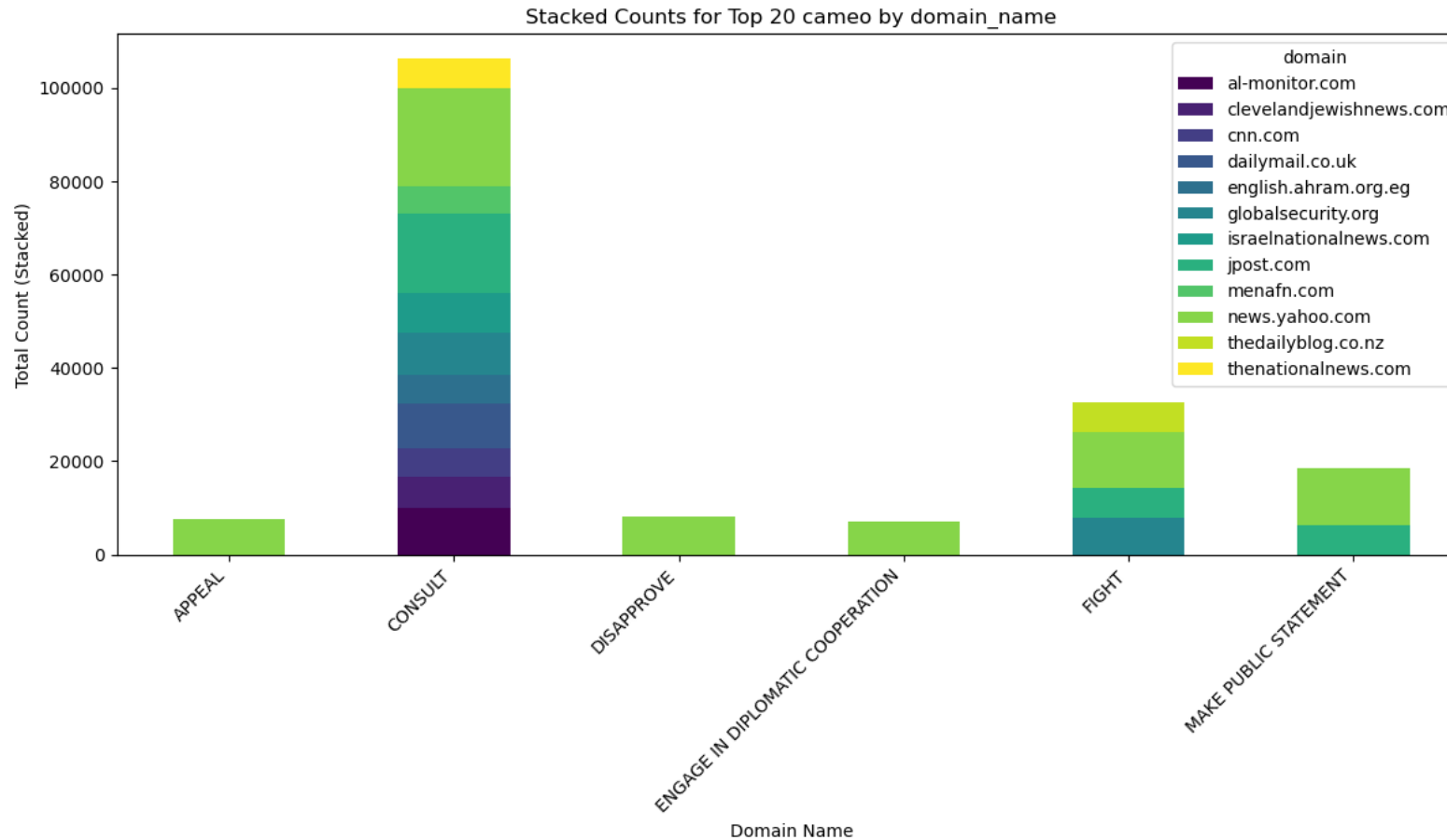


Figure 6: stacked counts for top 20 cameo by domain name

# We want to know how main media reports through GDELT

**Fastest-growing websites year-on-year among top 50 English language sites by global visits, March 2025**

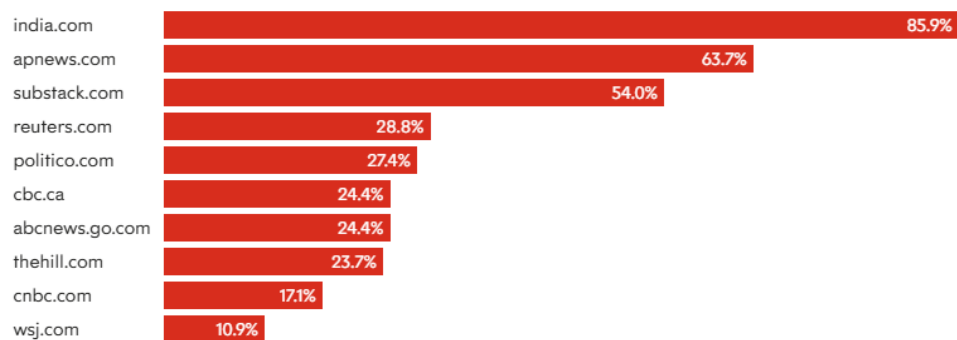


Chart: Press Gazette • Source: Similarweb

PressGazette

*Figure 8: Fastest growing news websites*

**Ten most-visited English-language websites in the world, change year-on-year, March 2025**



Chart: Press Gazette • Source: Similarweb

PressGazette

*Figure 9: The most visited English language website*

# Report distribution through cameo codes

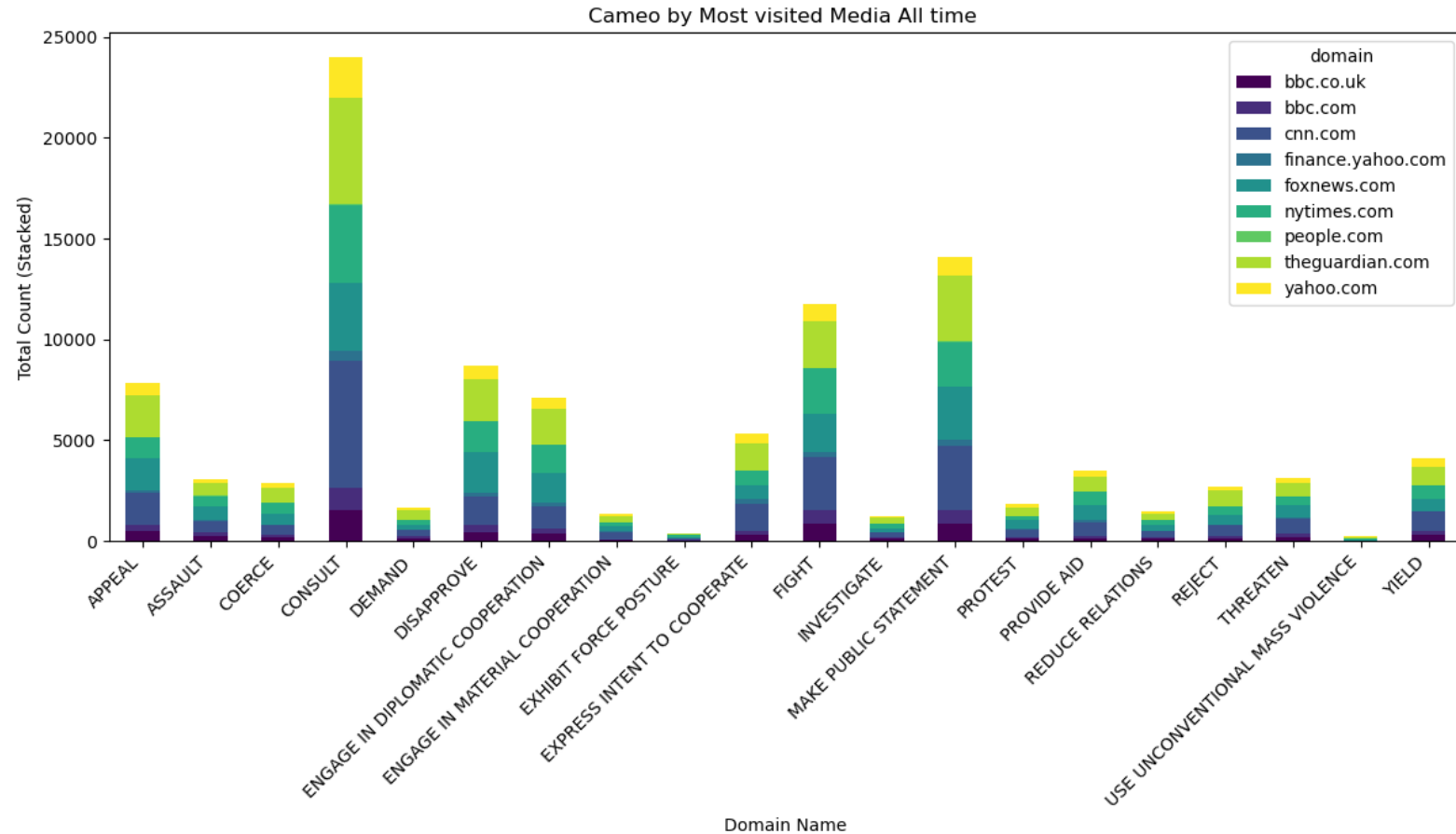


Figure 10: Cameo by most visited media all time



# Describing the Data: Actors

Israel the most frequent actor in news articles. Is Israel more active or more reported?

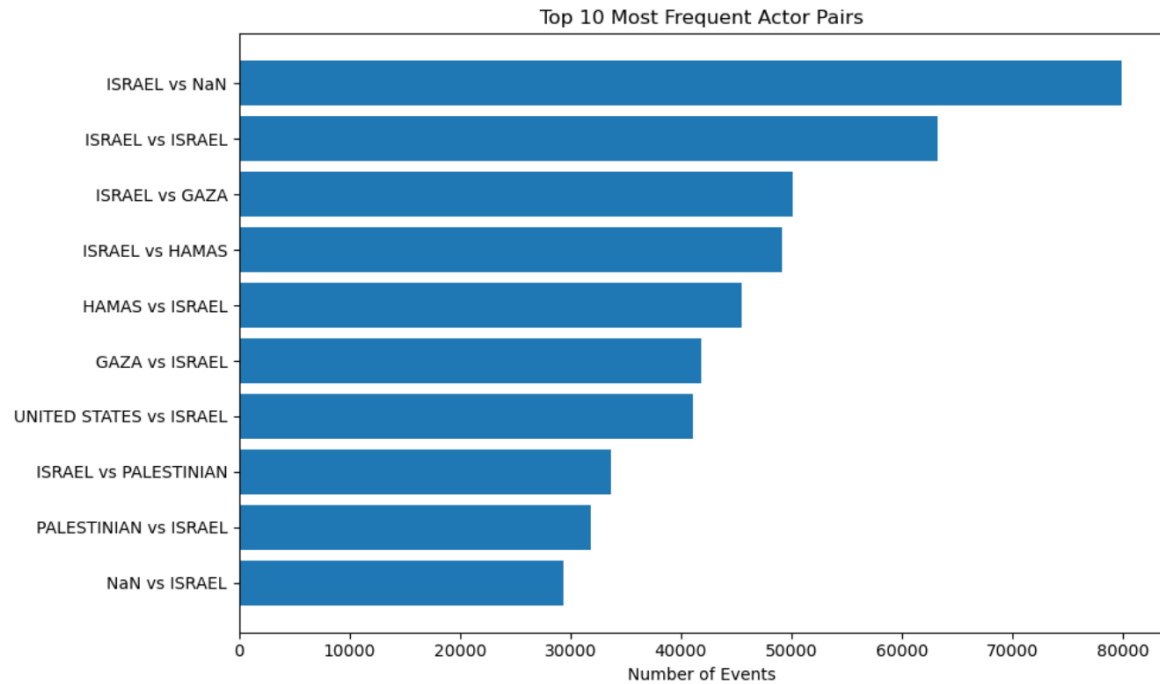


Figure 12: Top 10 most frequent actor pairs

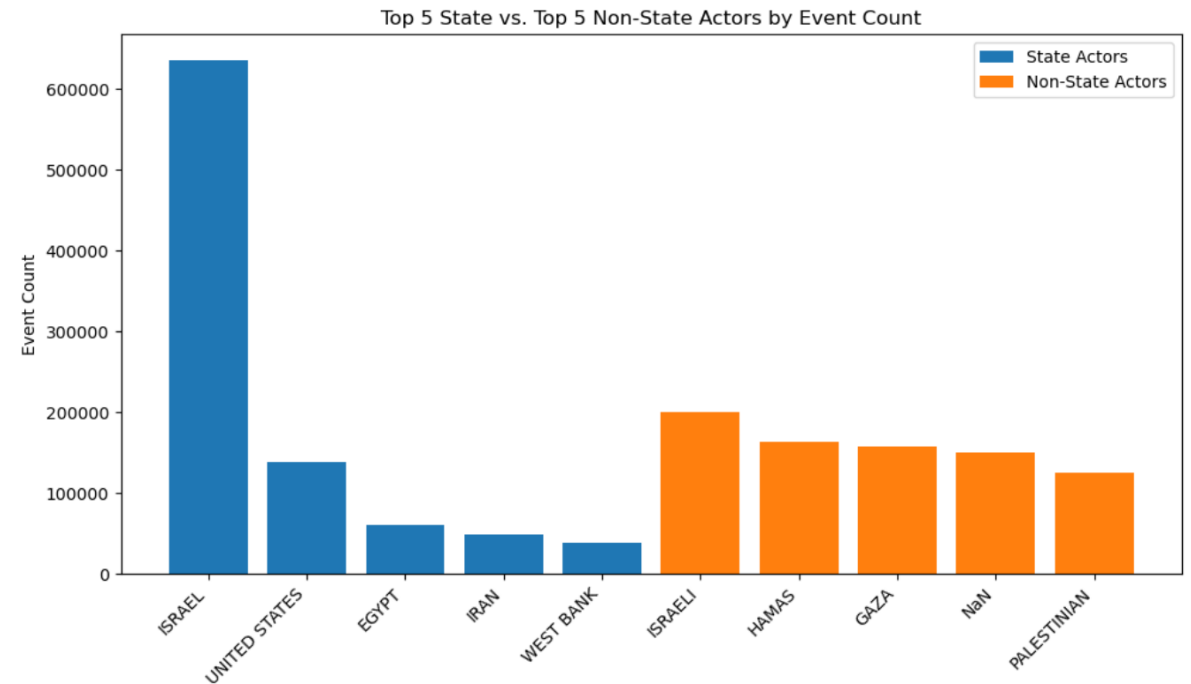


Figure 13: Top 5 state and non-state actor

# Polarization: By Event Type

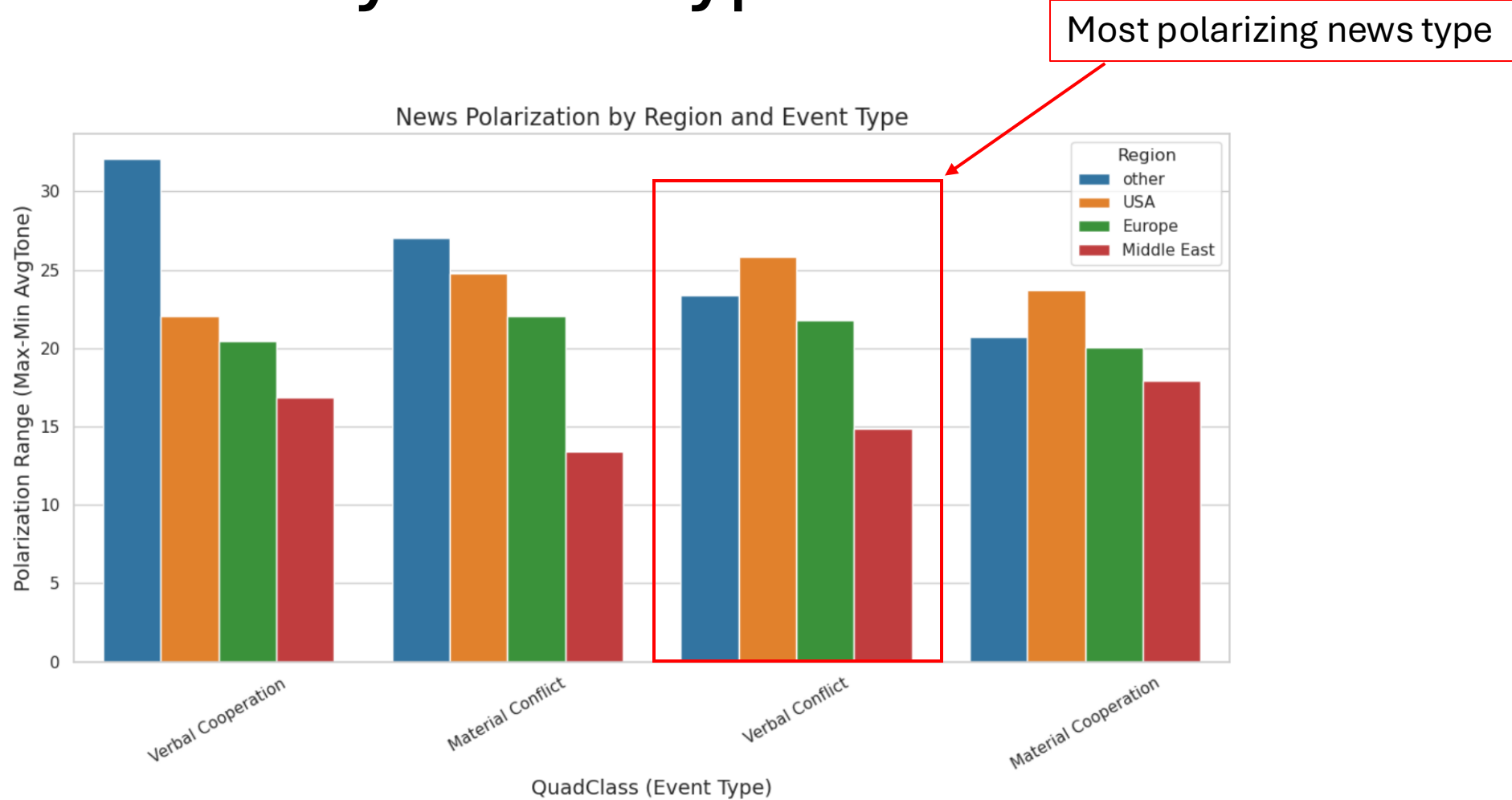


Figure 14: News polarization by region and event type

# Tone by Region

Material conflict between Israel and Palestine is covered more negatively in the Middle East compared to Europe and USA

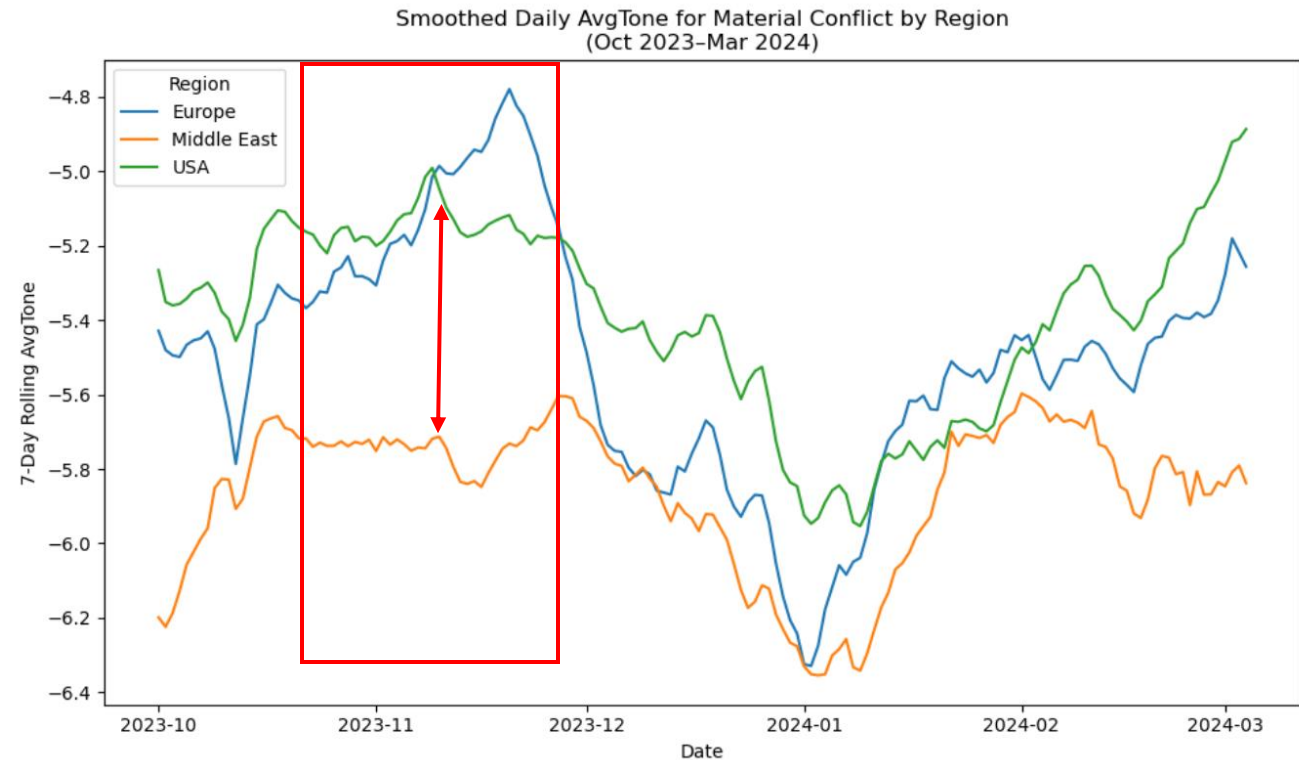
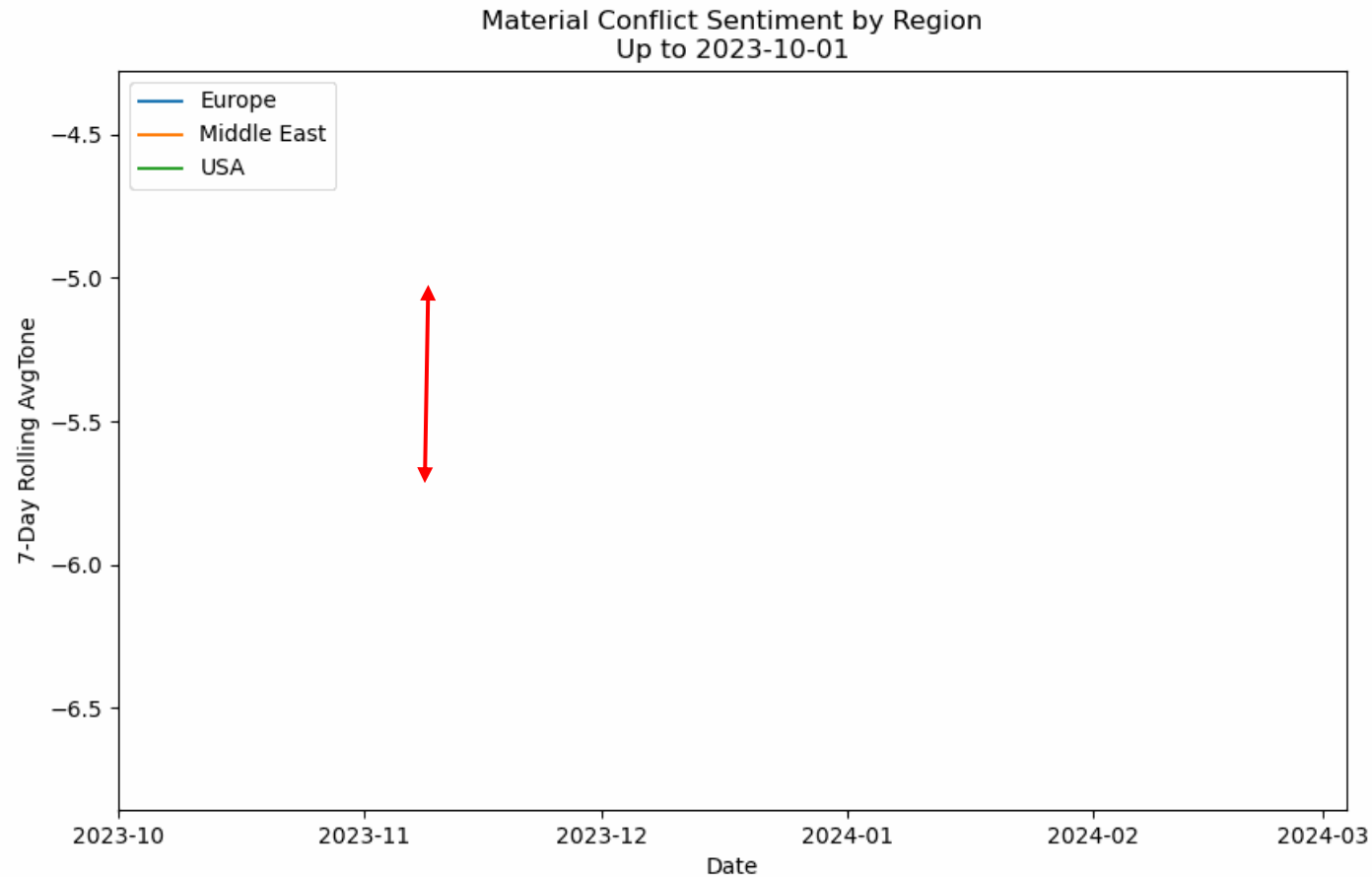


Figure 15: Average tone for material conflict covered in different region

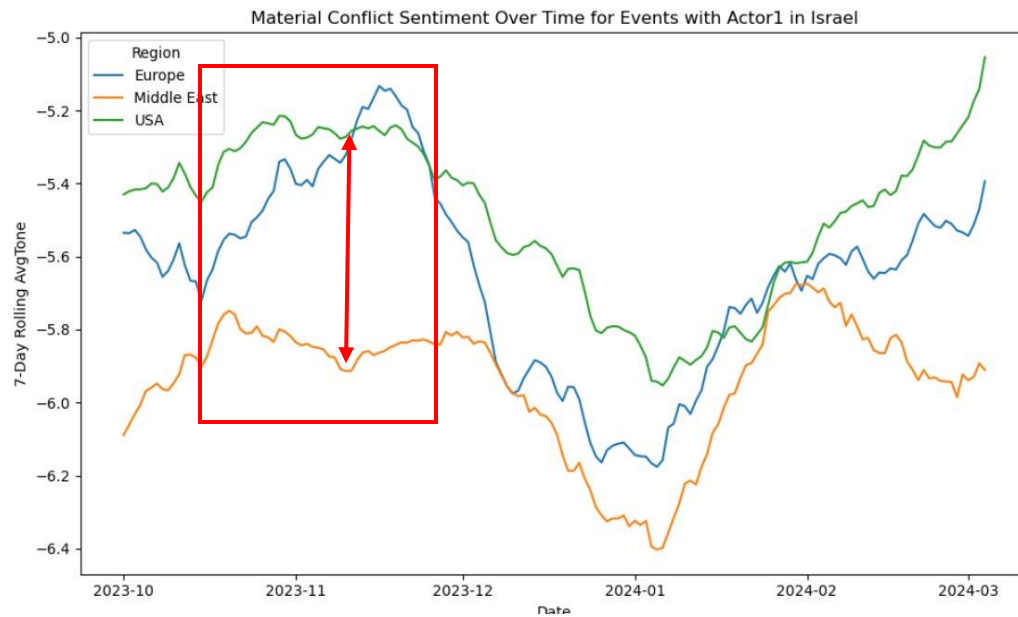


# Polarization: Sentiment for Material Conflict

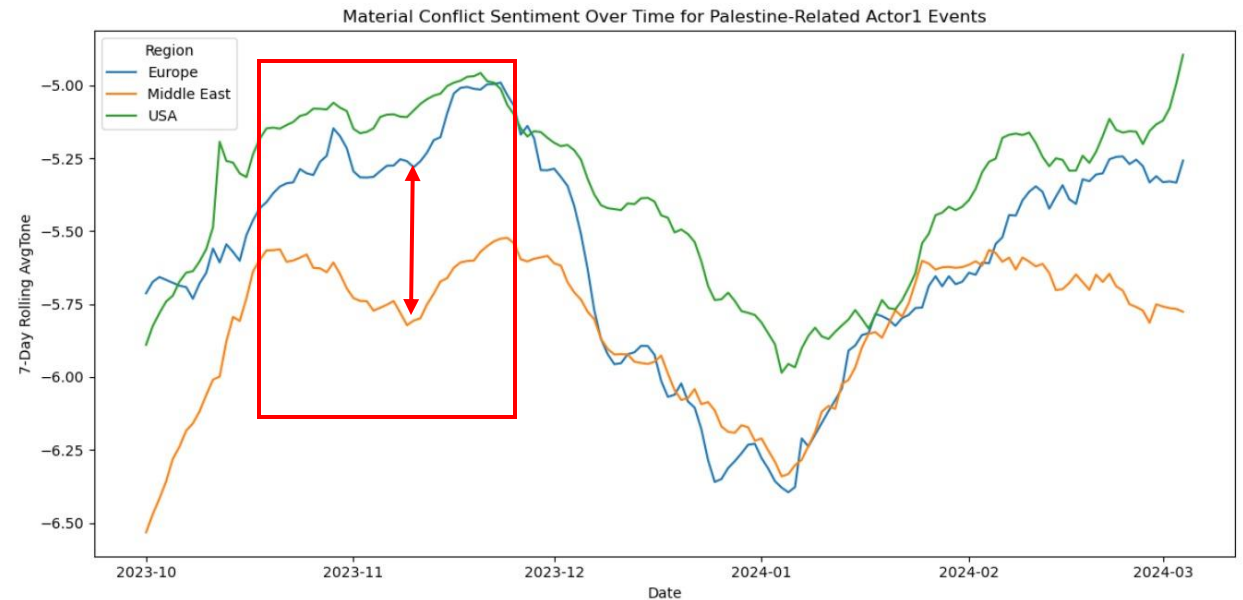


*Figure 16: Sentiment for material conflict in different region*

# Polarization: Reporting of Israel vs Palestine



*Figure 17: Regional sentiment for material conflict when committed by Israel*



*Figure 18: Regional sentiment for material conflict committed by Palestine*

# Top News Sources

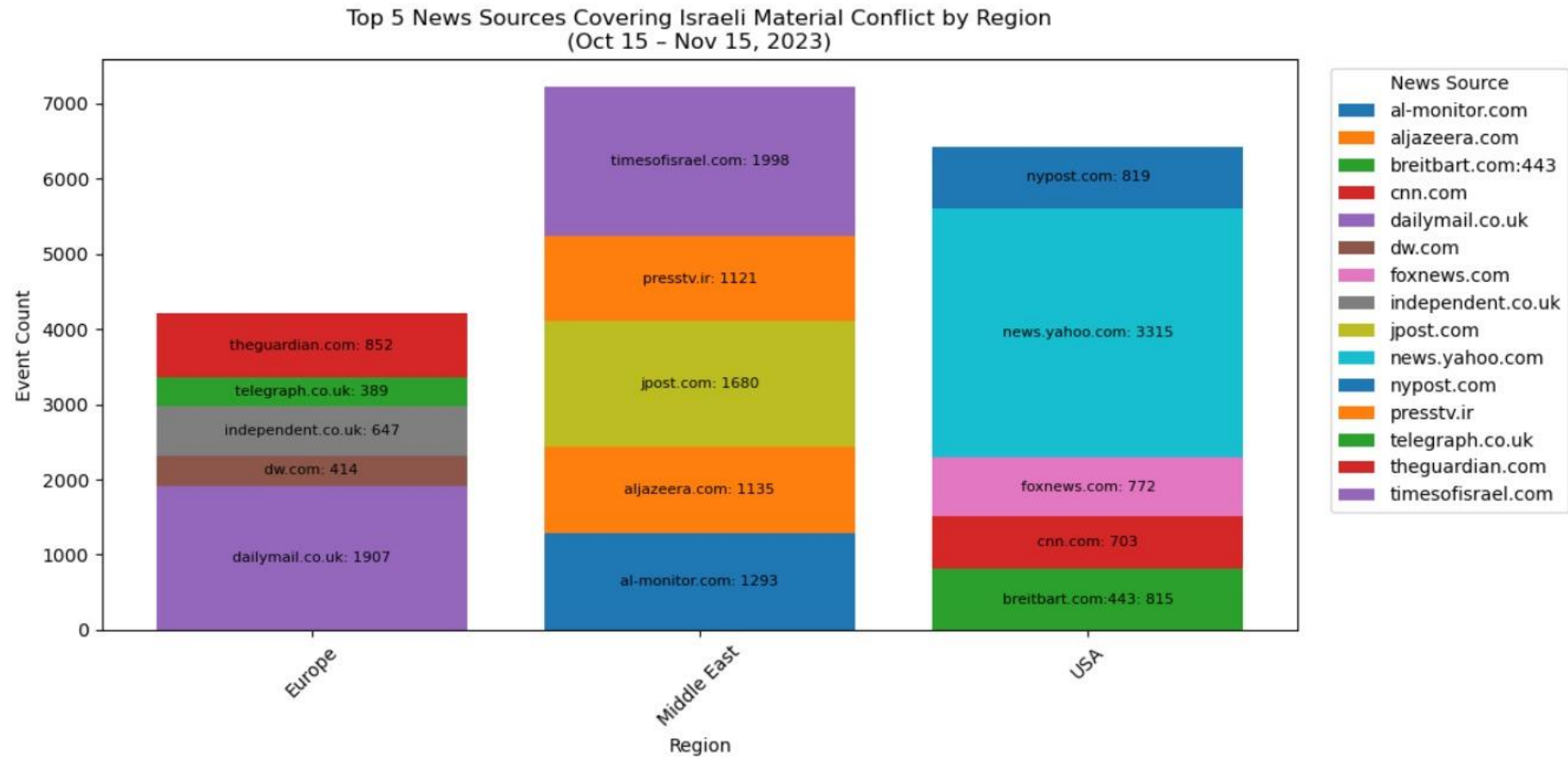


Figure 19: Top news sources that covered Palestine-Israel conflict



# Tone of Event Type per News Source

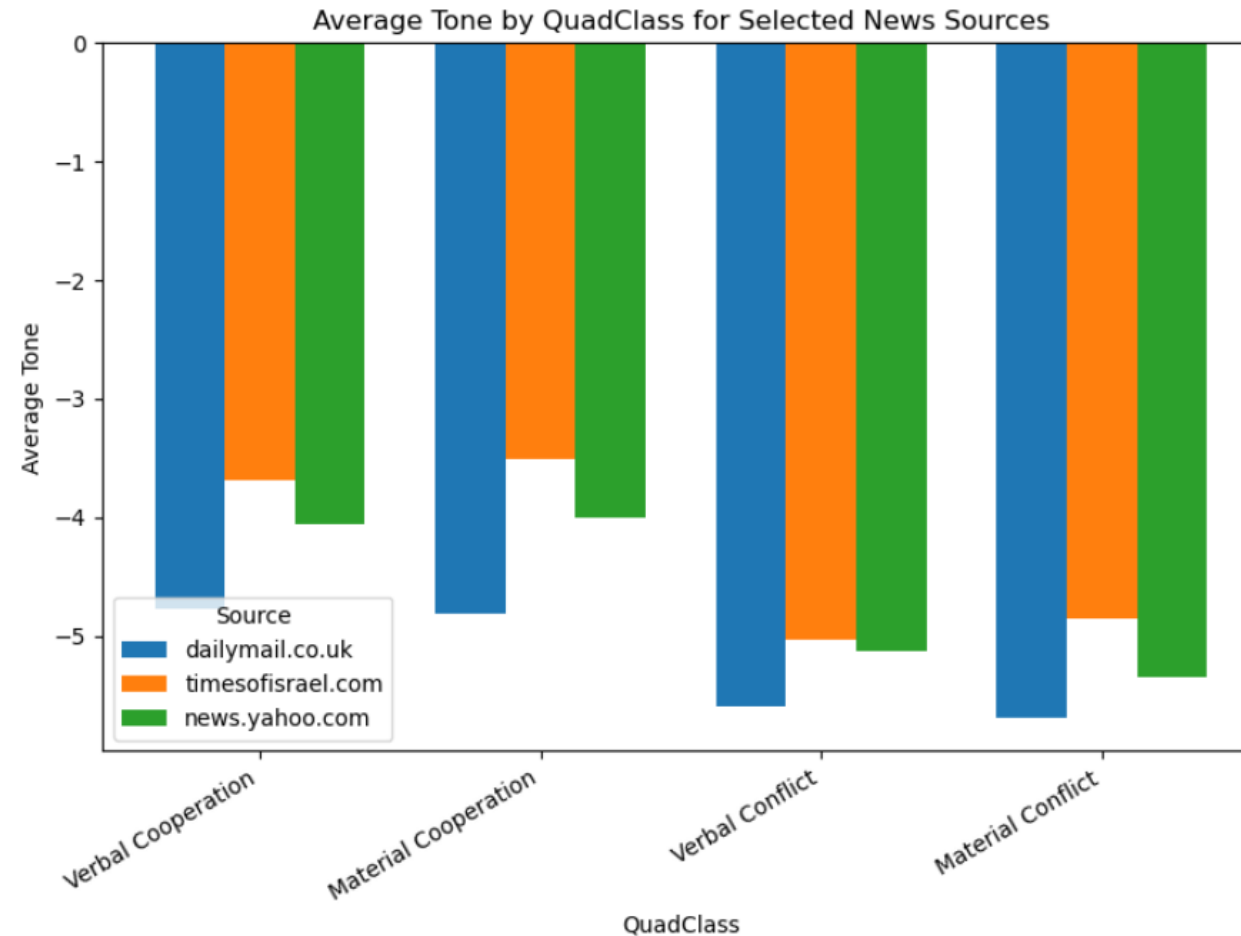


Figure 20: Average tone for each quadclass for top three news site

# Tone of Actors by Source

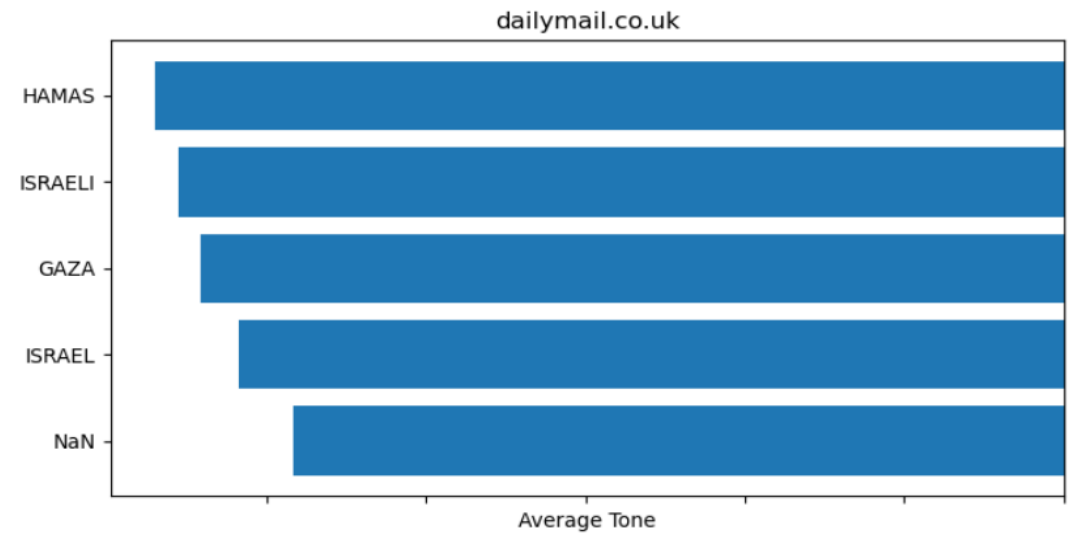
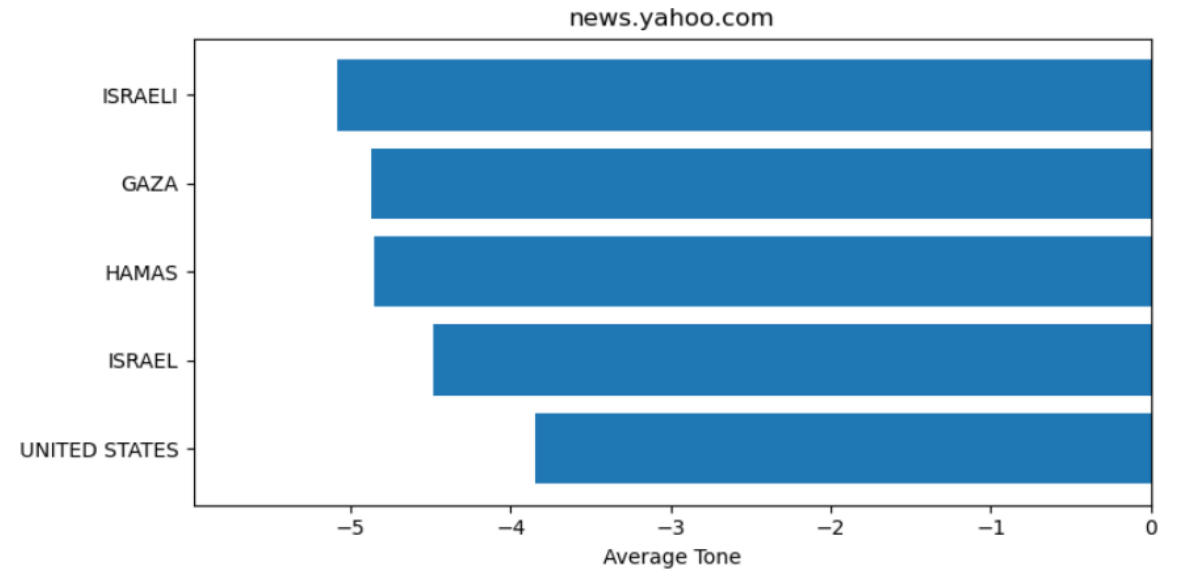
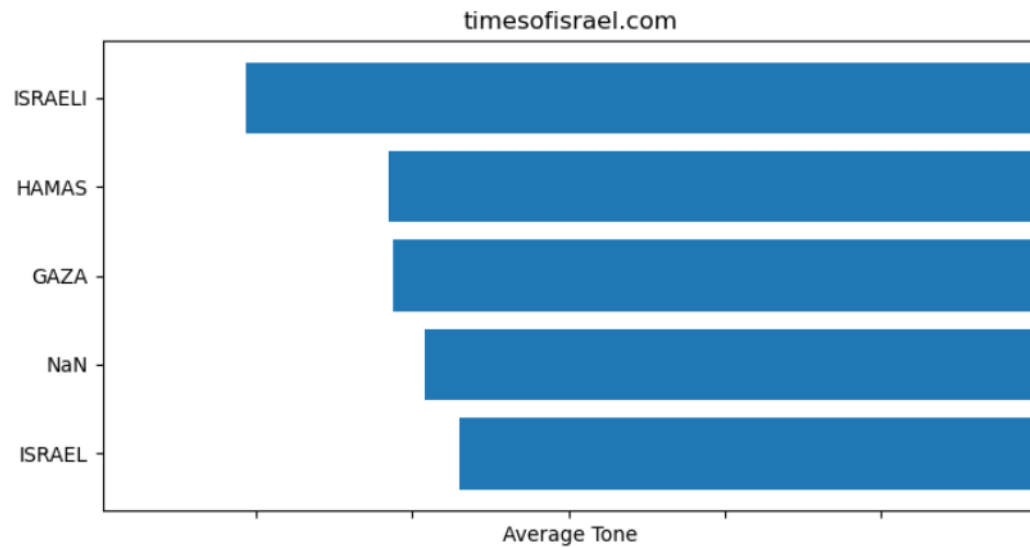


Figure 21: Top actors in top news sources

# Limitations

- Most news are from English speaking countries
- Are all the news about this conflict?
- GDELT tells about news coverage on events, not actual events



# Open Questions

- Are there differences in how differently affiliated media reports?
- What could be found out with GDELT 2.0 dataset?
- Questions from the audience



# References

## Background literature:

- Philip A. Schrodtt. (2012). CAMEO: Conflict and Mediation Event Observations Event and Actor Codebook
- Nikolov, D., Lalmas, M., Flammini, A., & Menczer, F. (2018). Quantifying Biases in Online Information Exposure. *Journal of the Association for Information Science and Technology*, 70. <https://doi.org/10.1002/asi.24121>.

## Images used in the presentation:

- Microsoft stock images
- Flickr user fabcom: fabcom\_DSC\_1054
- Flickr user scottgunn: Palestinian flag
- Flickr user del mich: сапунени мехурчета
- GDELT website