

# Research Questions

Geopolitical events are reported and interpreted through multiple information channels, including national news media and online social platforms. While prior computational studies have examined media coverage and social media discourse independently, less is known about how **the same real-world events** are narratively constructed across **different national media systems** and how these narratives compare to **public discourse on social media**.

In this study, we focus on U.S.–Venezuela relations between 2013 and 2025 and investigate narrative differences across media types and platforms using event-aligned news data and Reddit discussions. Specifically, we address the following research questions:

## RQ1. Cross-National Media Narratives

Do U.S. and Venezuelan news media construct different narratives when reporting on the **same geopolitical events** involving the United States and Venezuela?

## RQ2. News vs. Social Media Narratives

How do narratives in social media discourse (Reddit) differ from those in traditional news media when discussing the same events?

## RQ3. Cross-Platform Narrative Alignment

To what extent do social media narratives mirror, amplify, or diverge from national news media narratives over time?

## RQ4. Temporal Dynamics of Narrative Change

How do narrative differences across media systems and platforms evolve around major geopolitical events?

These questions motivate a comparative, event-centric analysis that aligns news coverage and social media discourse using GDELT event representations and large-scale Reddit data collected via the Arctic Shift API.

# Related Work

## Overview

Understanding how the same real-world geopolitical events are represented across different communication systems is a core topic in computational social science and natural language processing (NLP). Differences in narrative framing between national media and social platforms can reveal biases, agendas, and public discourse dynamics. Prior work has leveraged large-scale open event datasets like GDELT to study international relations and used social media archives such as Reddit to analyze how events are discussed publicly. However, limited work systematically compares **national media narratives for the same event** and contrasts these with **online public discourse**. Our project investigates narratives about U.S.–Venezuela relations (2013–2025) from U.S. news, Venezuelan news, and Reddit discussions, addressing how narratives align or diverge across media systems and platforms.

---

## Kwak & An (2016): Comparing Global News Event Databases (ICWSM)

**Summary.** This study conducts a comparative evaluation of major global news datasets, including GDELT, to examine properties such as coverage, scale, and event representation.

**Approach.** The authors analyze dataset characteristics and report differences in event counts, geographic coverage, and coding consistency.

**Main Findings.** Different news databases produce meaningfully different event signals, suggesting that quantitative studies of international relations should carefully select and validate the datasets used.

**Limitations.** The work focuses on dataset-level properties and does not analyze narrative framing or cross-national differences in how the same events are portrayed.

**Relevance.** Our project uses GDELT as the primary event representation source. This work motivates explicit filtering, source selection, and query documentation when using GDELT for narrative analysis.

---

## Olteanu et al. (2015): Event Coverage Gap Between News and Social Media (ICWSM)

**Summary.** This work investigates how major events are covered across online news and social media platforms, identifying systematic differences in attention and salience.

**Approach.** Using event-centric aggregation and time-series analysis, the study measures

coverage frequency, temporal patterns, and overlaps between news outlets and social media posts.

**Main Findings.** Not all events receive equal attention across media types; social platforms often lag or lead news in event attention, and patterns vary by event type.

**Limitations.** The study emphasizes coverage patterns rather than narrative framing. It does not analyze how events are discussed or described in text beyond frequency and timing.

**Relevance.** This work supports our second research question, which compares how Reddit discussions align or diverge from news narratives for the same events.

---

## Zhao et al. (2024): Event-Centric Framing and Media Attitude (EMNLP)

**Summary.** This paper proposes methods for detecting framing and media stance by leveraging structured event representations as an intermediary layer between raw text and higher-level narrative attributes.

**Approach.** The authors design event-aware framing models that use event participants and actions as anchors to extract narrative structure and media attitude.

**Main Findings.** Event-aware approaches improve interpretability and consistency when extracting framing compared to surface-level sentiment analysis alone.

**Limitations.** While demonstrating frame extraction efficacy, the methodology may need adaptation to cross-lingual and cross-national comparisons, especially where linguistic and journalistic norms differ significantly.

**Relevance.** Our project uses GDELT events as anchors to compare narrative frames between U.S. and Venezuelan outlets, aligning with the premise that events are a useful scaffolding for narrative extraction.

---

## Kuila et al. (2024): Aspect-Level Media Bias and Narrative Variance (ACM)

**Summary.** This study introduces computational models for measuring narrative variance and media bias by analyzing how outlets emphasize different aspects of the same topic.

**Approach.** The paper employs aspect-level modeling to quantify how different sources differentially emphasize coverage facets and frames.

**Main Findings.** Narrative differences can be measured as systematic divergence in aspect distributions, offering a fine-grained view of media bias beyond sentiment.

**Limitations.** Aspect-based approaches require careful domain design; identifying appropriate aspects for geopolitical events may require expert knowledge or manual annotation frameworks.

**Relevance.** Our analysis operationalizes narrative divergence in part through frame/aspect comparisons of event-aligned article sets from different national media sources.

---

## Arctic Shift API for Reddit Collection (Resource Paper / Tool Reference)

**Summary.** The Arctic Shift API provides a research-oriented interface for accessing historical Reddit submissions and comments, which has been used in recent computational research to support large-scale social media analysis.

**Approach.** Programmatic retrieval of Reddit text with time windows, supporting long-running datasets suitable for longitudinal narrative studies.

**Main Findings.** Arctic Shift enables comprehensive Reddit data collection while logging metadata about collection parameters and scope, alleviating impacts from platform limitations.

**Limitations.** Reddit archives may suffer from deletion biases, and API query constraints can affect coverage completeness.

**Relevance.** Our project's Reddit dataset was acquired using Arctic Shift. This work supports documentary evidence for the chosen API tool and data collection methodology.

---

## Research Gap and Contribution

Collectively, these works demonstrate that (1) global news event databases like GDELT can provide large-scale, structured event signals but may vary by dataset properties, (2) social media and news platforms show systematic differences in event attention and timing, and (3) narrative framing and bias can be computationally extracted at fine granularity. However, **few computational studies directly compare national media narratives for the same real-world events, and even fewer relate these differences to social media discourse dynamics across platforms.** By combining GDELT event alignment with Reddit narrative analysis, this project fills a gap at the intersection of event-centric framing and multi-platform narrative comparison.

# Datasets

## GDELT: Global Database of Events, Language, and Tone

- **Dataset(s) Used**  
GDELT 2.x event-level and document-level data covering the period **2013–2025**, focusing on events involving the United States and Venezuela.
  - **Why Selected**
    - **Event Alignment:** GDELT provides structured event representations (actor–action–actor, time, location), enabling alignment of *identical real-world events* across different national media systems.
    - **Scale and Coverage:** Global, multi-year coverage supports longitudinal analysis of international relations narratives.
    - **Domain Fit:** Event-centric representations allow narrative comparison conditioned on the same geopolitical events, which is central to our research questions.
  - **Public Availability**  
GDELT is a publicly available dataset.
  - **Official Link**  
<https://www.gdeltproject.org/data.html>
  - **Downloaded Data Location**  
For reproducibility and offline analysis, we downloaded and stored the relevant GDELT data locally. The processed and organized data are available at the following shared location:  
**Google Drive (GDELT folder)**  
[https://drive.google.com/drive/folders/1Ej5OVsdoTVQN3natA525Mo\\_qytrR5HX8](https://drive.google.com/drive/folders/1Ej5OVsdoTVQN3natA525Mo_qytrR5HX8)  
→ `gdelt/`
- 

## Reddit Discourse Data (Collected via Arctic Shift API)

- **Dataset(s) Used**  
Reddit submissions and comments related to U.S.–Venezuela events from **2013–2025**, collected using the **Arctic Shift API**.
- **Why Selected**
  - **Complementary Narrative Source:** Reddit provides user-generated discourse that reflects public interpretation, reaction, and reframing of geopolitical events reported in the news.
  - **Temporal Alignment:** The data span aligns with the GDELT time window, enabling event-level and temporal cross-platform comparisons.

- **Granularity and Metadata:** Post- and comment-level text, timestamps, and subreddit information support narrative, temporal, and community-level analyses.
- **Public Availability**

Raw Reddit data redistribution is subject to platform and policy constraints. Therefore, the raw collected data are **not redistributed publicly**.
- **Data Collection Method**

Data were collected programmatically using the Arctic Shift API with predefined keyword and time-range filters relevant to U.S.–Venezuela relations.
- **Downloaded Data Location**

The collected Reddit data are stored locally and shared via Google Drive for project collaboration:

**Google Drive (Reddit folder)**  
[https://drive.google.com/drive/folders/1Ej5OVsdoTVQN3natA525Mo\\_qytrR5HX8](https://drive.google.com/drive/folders/1Ej5OVsdoTVQN3natA525Mo_qytrR5HX8)  
→ reddit/

---

## Code and Reproducibility

- **GitHub Repository**  
<https://github.com/Shinhunjun/multiplatform-war-narratives>
- **Repository Contents**
  - Data collection scripts (GDELT queries, Arctic Shift API usage)
  - Preprocessing and analysis pipelines
  - Derived datasets and aggregated features (no raw Reddit dumps)
- **Note on Data Sharing**

Due to size constraints and platform redistribution policies, raw datasets are stored externally (Google Drive), while the GitHub repository provides all scripts and derived outputs necessary to reproduce the analysis.