

A1. Data Sources

The dataset consists of two folders of news articles collected to analyze media sentiment and framing across two geopolitical events: The invasion of Venezuela and the Trump brokered ceasefire deal in Israel-Gaza.

Articles were obtained from publicly and subscription available online news sources, including:

- BBC
- CNN
- Fox News
- CBS News
- AP News
- PBS
- USA Today
- New York Times
- NBC
- ABC
- MSNBC
- NPR
- Wall Street Journal
- Washington Post
- Forbes
- New York Post

For each story, articles were selected based on relevance to the primary topic using keyword searches such as:

- **Invasion of Venezuela:** “Invasion”, “Venezuela”
- **Trump Ceasefire Deal in Israel-Gaza:** “Trump”, “Ceasefire”, “Israel-Gaza”

A2. Data Collection

Each article was downloaded as a plain text (.txt) file and stored in separate folders by story:

- /Venezuela/
- /Israel/

Each file represents a single news article.

A3. Text Pre-Processing

Prior to analysis, text files were processed within a TextEditor. The following steps were applied:

1. Copy and paste news articles piece by piece to exclude advertisements and other non-news sections
2. Change entire file to plain text to remove hyperlinks
3. Ensure all text is plain for input into model

A4. Sentiment Analysis Preparation

Text was prepared for sentiment analysis using a spaCy scoring tool

For each article, the following metrics were generated:

- Positive sentiment score
- Negative sentiment score
- Neutral sentiment score

These scores were later aggregated by news organizations for comparative analysis.

A5. Reproducibility

All preprocessing and analysis were conducted using Python (version 3.14.3) with the following libraries:

- Pandas / Numpy
- spaCy
- Hugging face
- matplotlib / seaborn