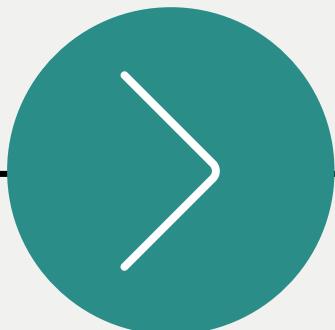


The Israeli-Palestinian conflict:

# A sentiment and linguistic analysis



Pasquale Caponio - Enrico Cipolla Cipolla - Luca Soleri - Ossama Tchina

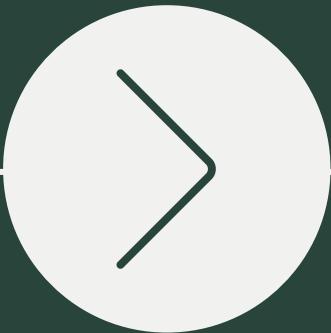




# Table Of Content

- Introduction
- Background
- Data Acquisition
- Data Analysis
- Results
- Conclusions

# Introduction



**What are the linguistic variations  
and sentiment trends  
surrounding the Israel-Palestine  
conflict across different media  
platforms?**

# Objectives



Analyze the sentiment surrounding the conflict and how it changes over time

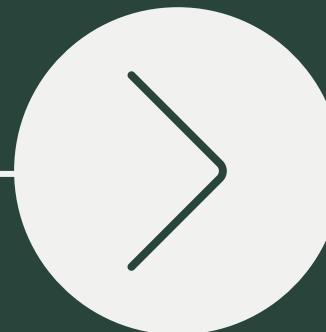


Identify differences within public and social media in the language they use to convey information



Understand how the associations between controversial words evolved as the conflict developed

# Background



# What is happening?

Starting point: 7th October 2023, Hamas organized an attack on the Southern of Israel



Today:	Killed	Injured
Palestine	+34,000	+76,000
Israel	+1,200	+5,400

# Data Acquisition



# Sources of Data

## Social Media

- Reddit



reddit

## Newspapers

- The New York Times (USA)
- The Guardian (UK)
- Al Jazeera (Qatar)

The  
New York  
Times

The  
Guardian

 ALJAZEERA

# Data from Reddit



## Description

Dataset found on Kaggle containing **2'000'000+** **comments** from posts related to the Israel - Palestine conflict

## Useful info

- Dataset is updated daily with new comments
- For each comment its timestamp is stored



# Data from NYTimes

## Description

Data scraped from the NYTimes. Final dataset containing **1500+ articles** about the Israel-Palestine conflict.

## Scraping Procedure:

### › Retrieving URLs

Using the NYTimes API, we stored all the articles' URLs related to the Israel - Palestine conflict

### › Generating Log-in cookies

We manually logged into our NYTimes account and extracted the cookies of this session



### › Scraping Articles

Using Selenium, we access each article page using the stored URLs list. We retrieve all paragraphs within the page



# Data from The Guardian

## Description

Data scraped from "The Guardian". Final dataset containing **300+ articles** about the Israel-Palestine conflict.

## Scraping Procedure:



### Retrieving URLs

Using the Guardian API, we stored all the articles' URLs related to the Israel - Palestine conflict



### Scraping Articles

Using requests & Beautiful Soup, we access each article page using the stored URLs list. We retrieve all paragraphs within the page



# Data from Al Jazeera

## Description

Data scraped from "Al Jazeera". Final dataset containing **2000+ articles** about the Israel-Palestine conflict.

## Scraping Procedure:



### Retrieving URLs

Using Selenium, we interacted with the website to retrieve all the articles' URLs related to the Israel - Palestine conflict



### Scraping Articles

Using requests & Beautiful Soup, we access each article page using the stored URLs list. We retrieve all paragraphs within the page

# Data Analysis



# Word2Vec



## Preprocessing Steps

Tokenizing, Stopwords  
Removal and Lemmatizing



## Motivation

Capturing the Semantic  
associations between debated  
words using w2v.similarity() function

## We analyzed

{Hamas, Terrorist}

{Israel, Selfdefence}

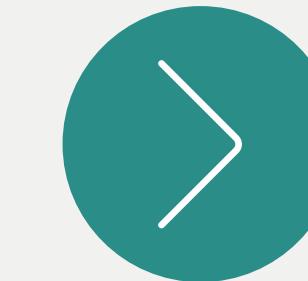
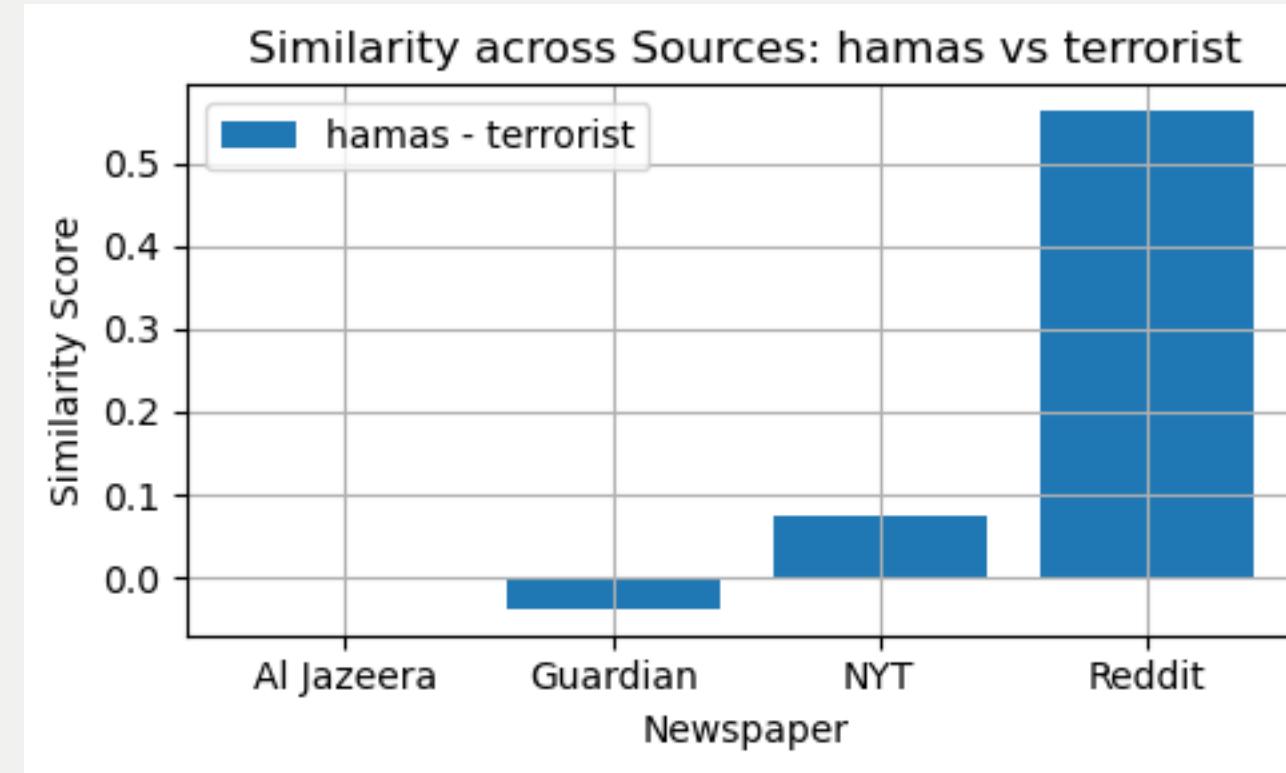
{Palestine, Selfdefence}

{Israel, Genocide}

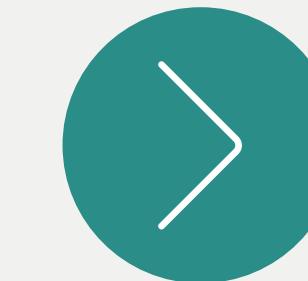
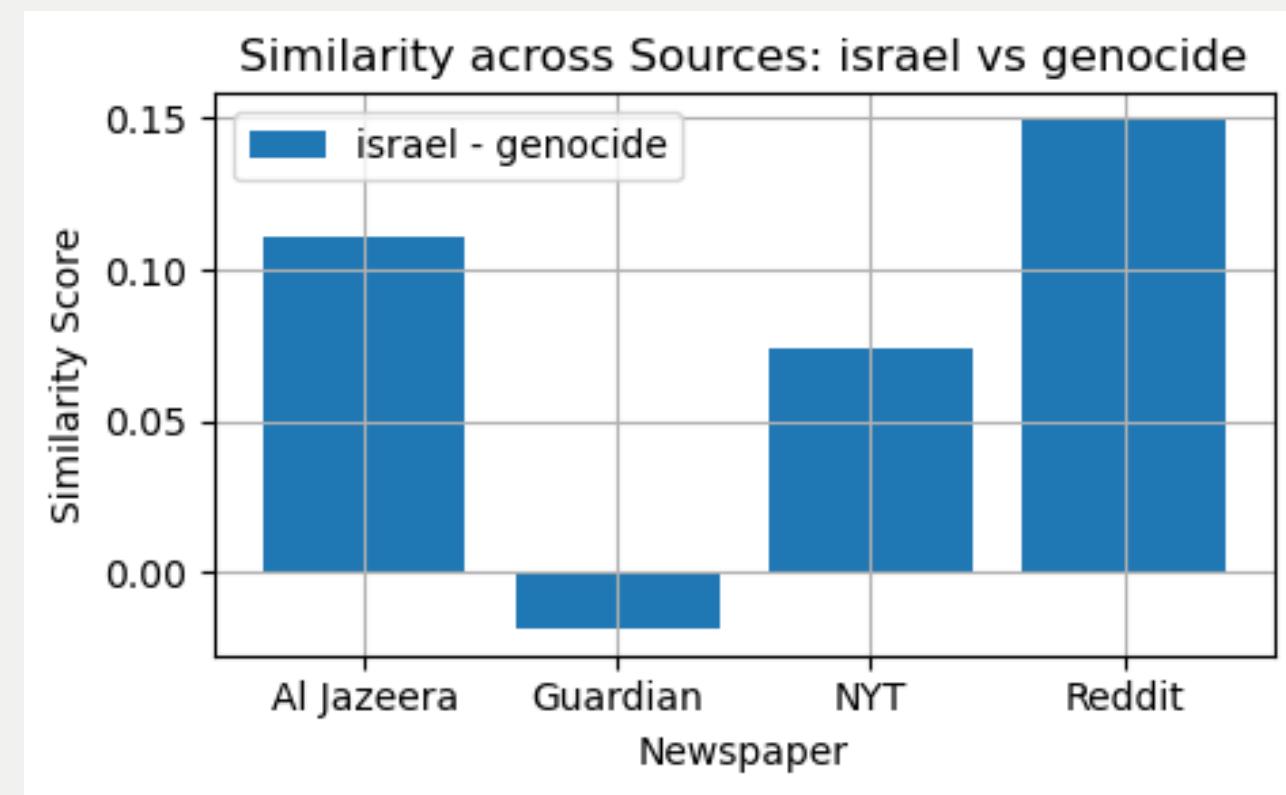
{Israel, Ceasefire}

```
Word2Vec(vector_size=100,  
         window=10,  
         sample=0.01,  
         epochs=50,  
         negative=10,  
         min_count=10,  
         workers=cpu_count(),  
         hs=0,
```

# Newspapers Comparative Analysis



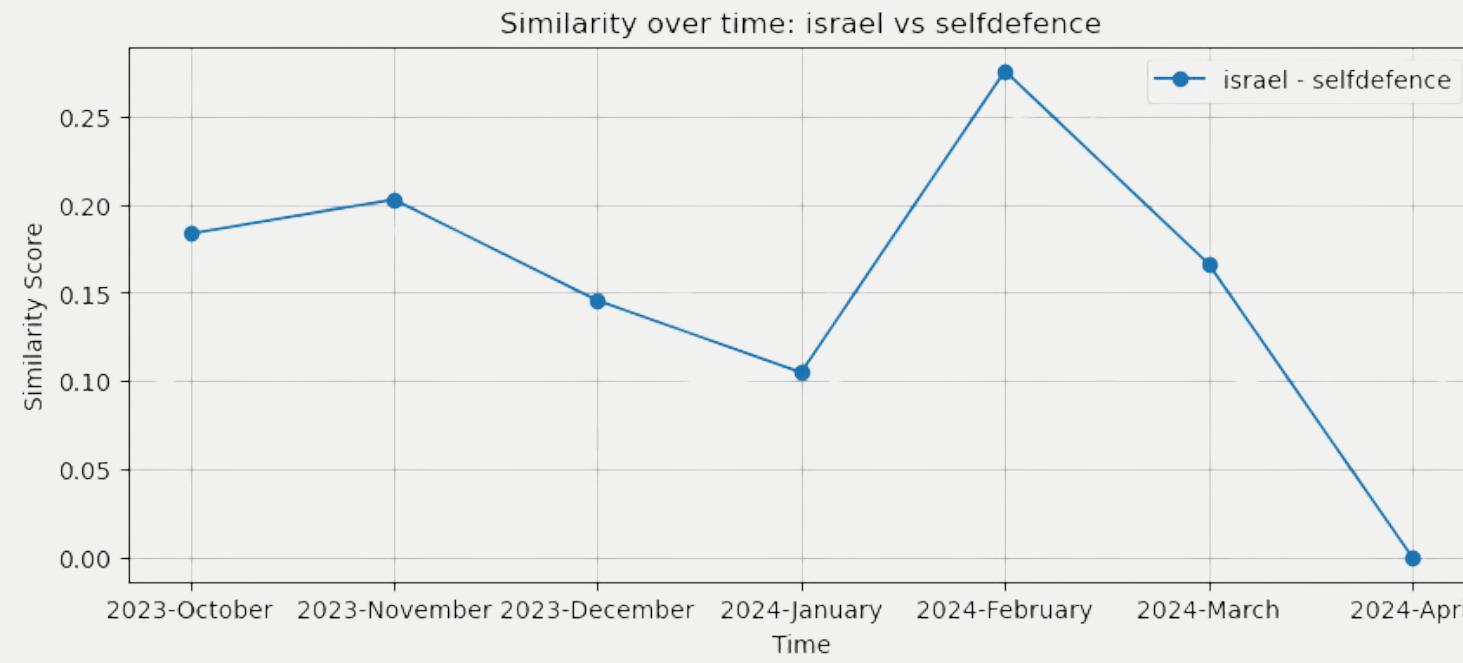
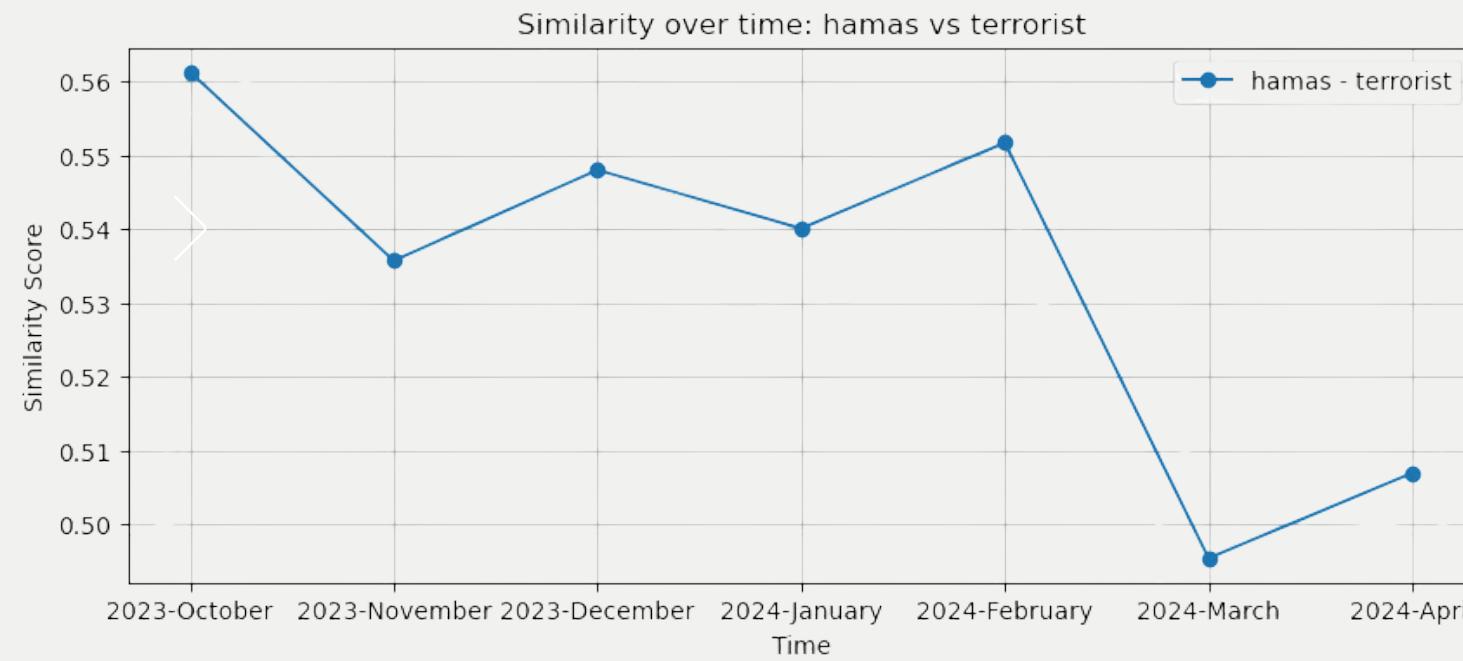
**Reddit has the highest similarity score, whereas for Al Jazeera the two values are not related at all.**



**Al Jazeera's value is quite close to the Reddit one: high critical positions towards Israeli actions**

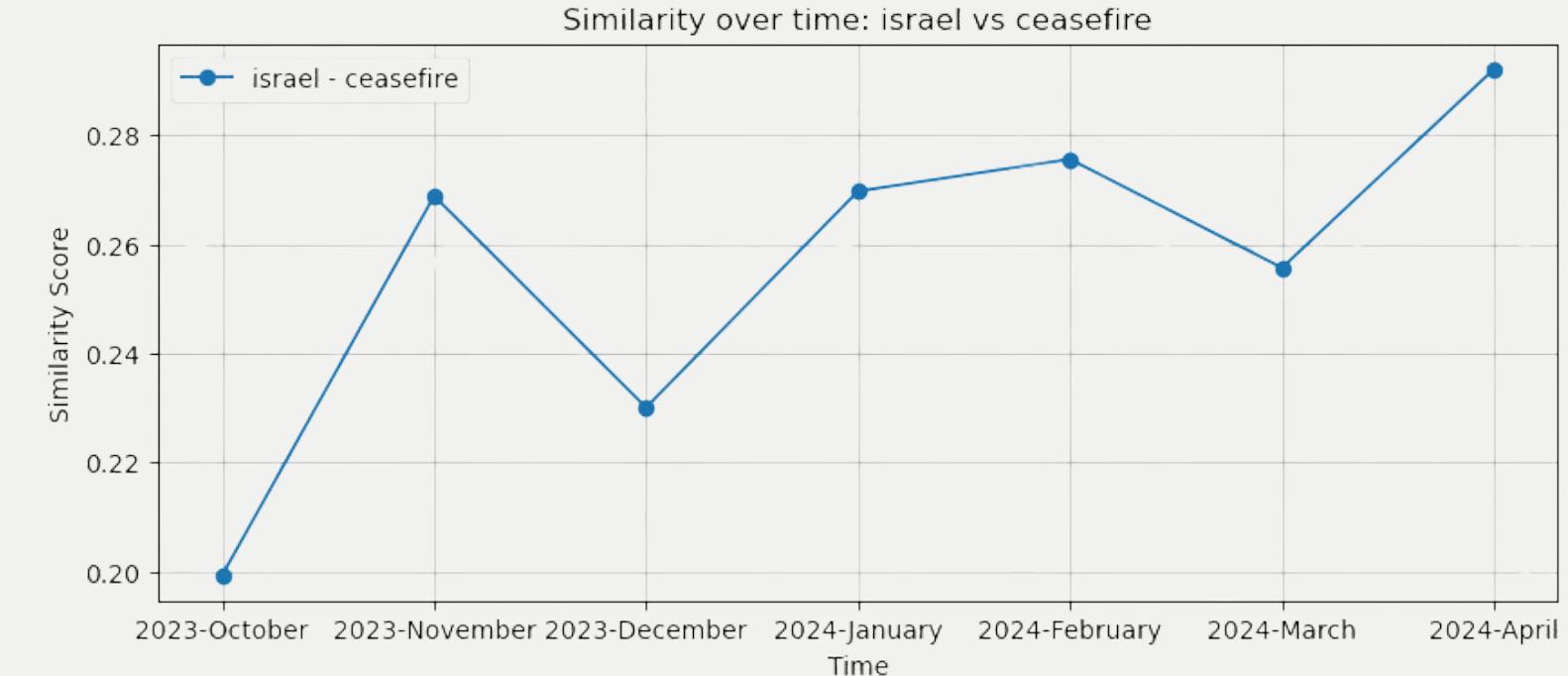
# Reddit Overtime Analysis

Hamas vs terrorist



Israel vs selfdefence

Israel vs ceasefire



Users are converging towards  
peace, rather than war

# Sentiment Analysis



## Motivation

Capture the evolution of the sentiment towards the war and frame key points in the conflict



## Preprocessing Steps

Tokenizing,  
Stopwords Removal  
and Lemmatizing to  
boost sentiment  
assessment  
accuracy



## Model

VADER sentiment  
analysis tool,  
calculating the  
sentiment score for  
each text



## Results

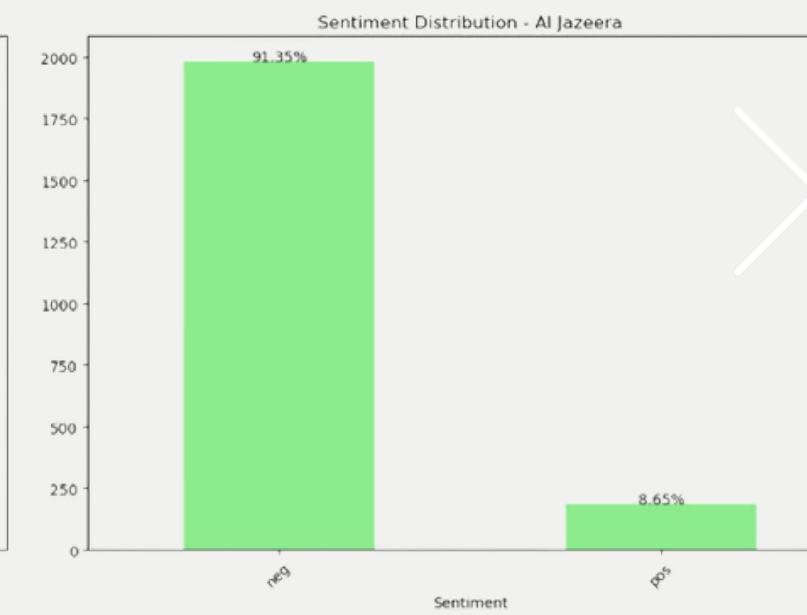
We obtained the overall sentiment and the evolution of the sentiment over time for all the datasets.

# Overall Sentiment Analysis

Reddit

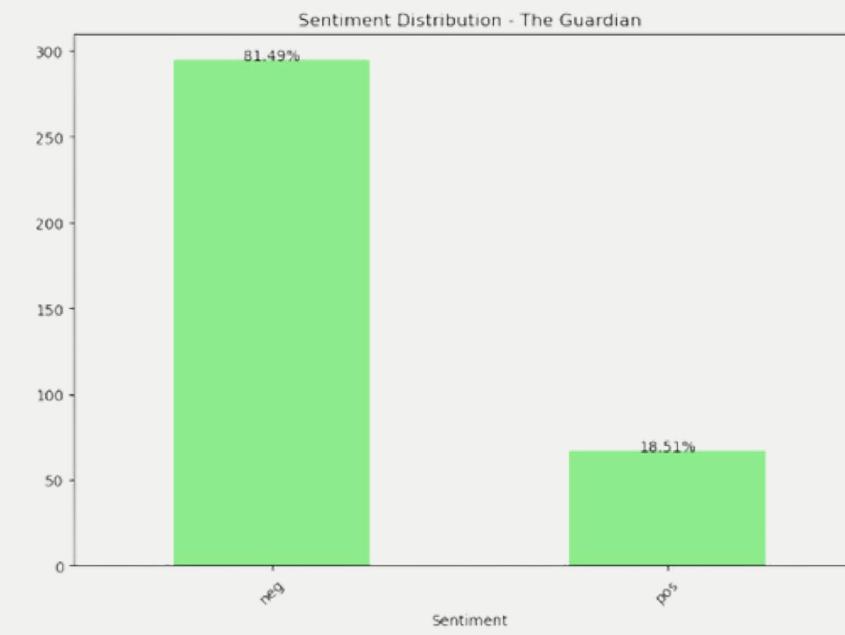


Al Jazeera

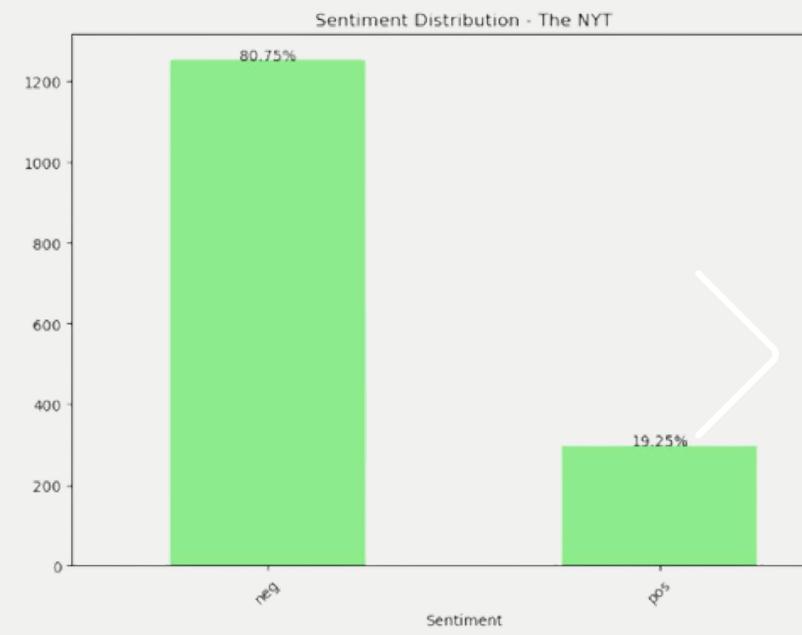


**The sentiment towards the conflict is predominantly negative**

**Reddit users express stronger and more controversial opinions compared to traditional media outlets**



The Guardian

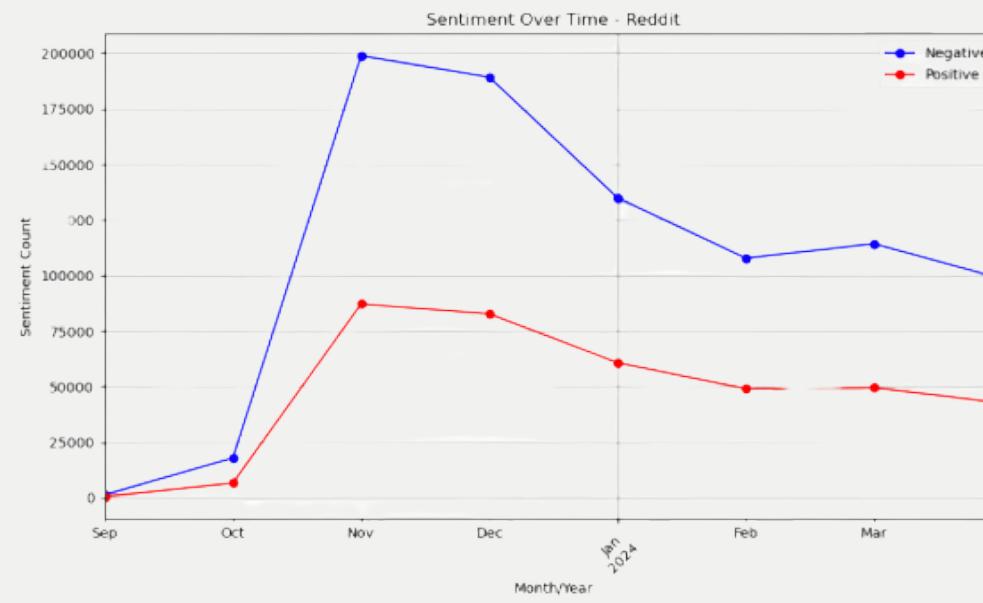


The NYTimes

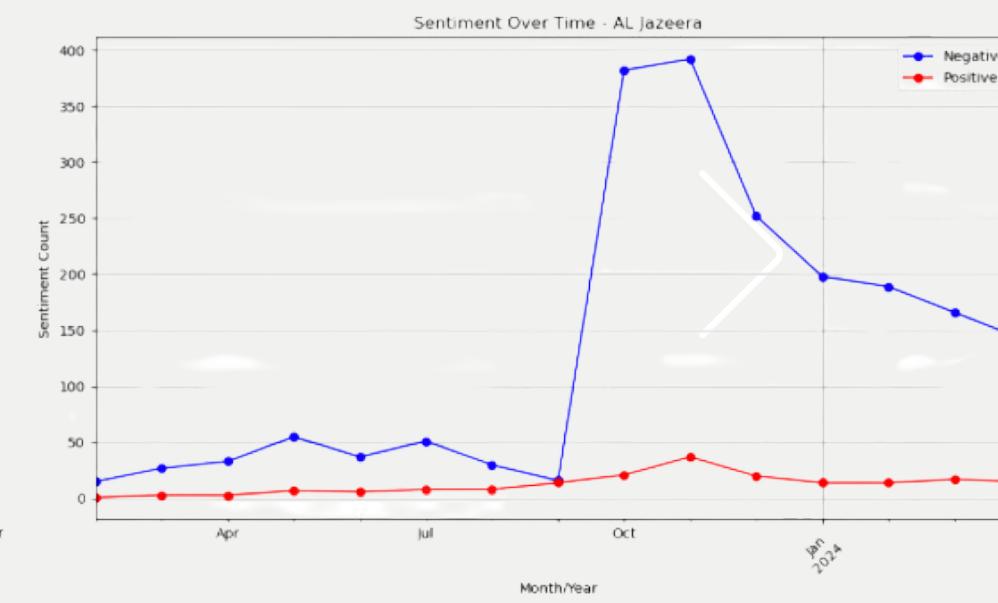
**Al Jazeera's exhibits the highest negativity towards the conflict**

# Sentiment Analysis Overtime

Reddit

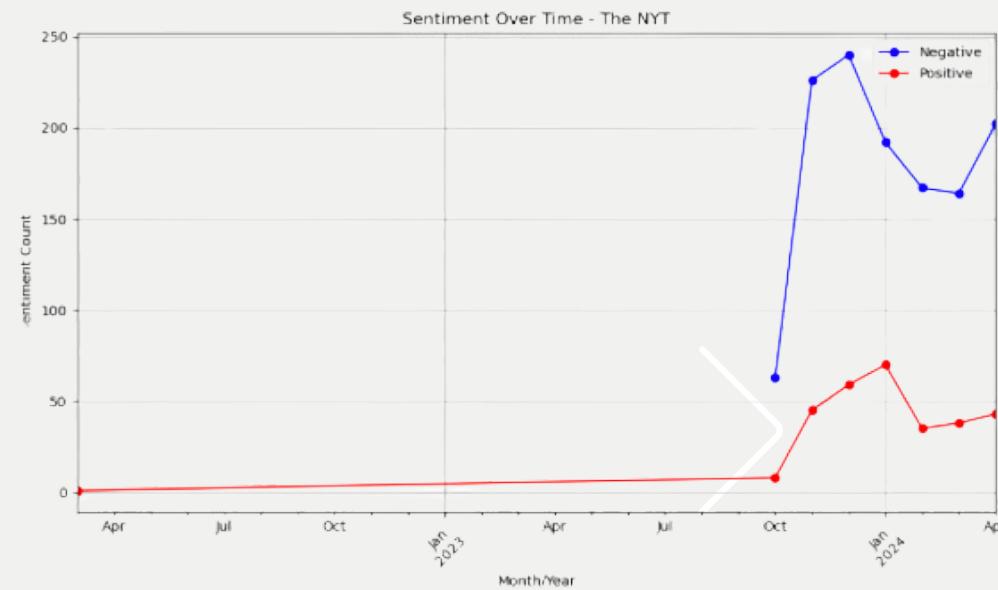
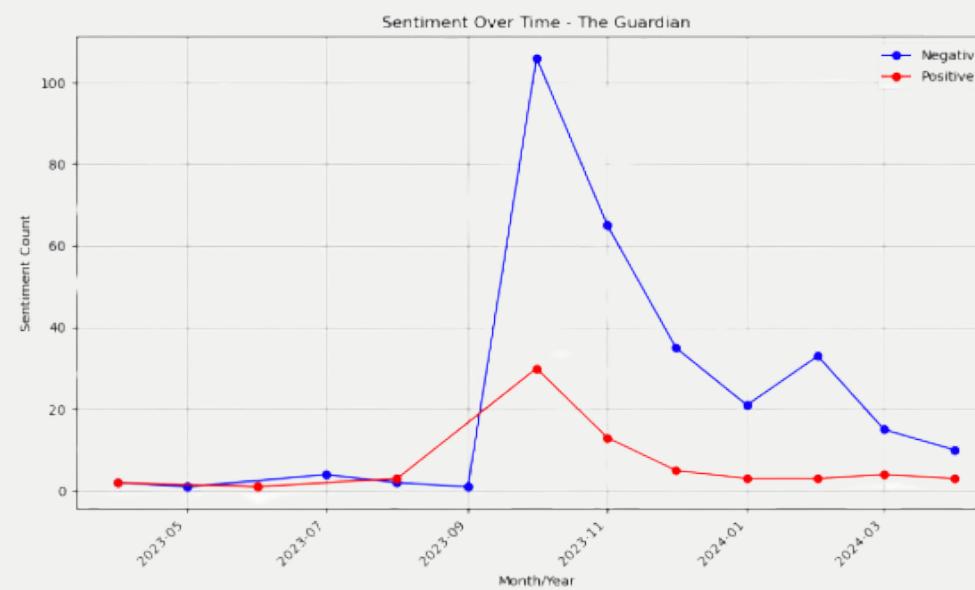


Al Jazeera



**Sentiment across all media outlets shows a negative spike in October during the conflict's onset**

**Western media, especially The Guardian, quickly see sentiment decline**

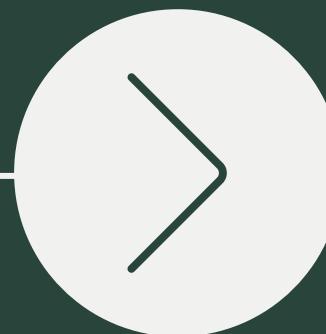


The Guardian

The NYTimes

**Increases in March and April, related to what a media decides to focus on**

# Conclusions



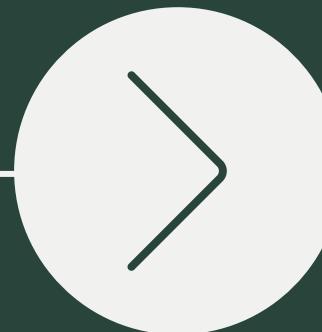
# Main Takeaways

We found **significant variability** in portraying key themes like "self-defense" and "aggression" across different platforms highlights **media biases**.

Compared to traditional media, Reddit shows more **direct and polarized responses** to the conflict.

Consistently **negative sentiments** across platforms intensify after major events, especially noted on Reddit and Al Jazeera.

# Any Questions?



**Thank you!**

