Digital Discourse vs. Real-World Events: A Computational Analysis of Reddit and Conflict Data in the 2025 Palestine War *

Anteneh Getachew Yitayal
anteneh.yitayal@studenti.unitn.it

Tereza Saskova
tereza.saskova@studenti.unitn.it

Sanzhar Sailaubek
sanzhar.sailaubek@studenti.unitn.it

July 18, 2025

Abstract

This study examines the relationship between Reddit discussion and real-world conflict events during the Spring 2025 ongoing conflict in Palestine. We explore how online engagement reflects—or fails to reflect—the real-life violence by comparing Reddit comments with conflict data from the Armed Conflict Location and Event Data Project (ACLED). Using a computational approach involving multiple methods, we apply Latent Dirichlet Allocation (LDA) to identify changing discussion topics and use a BERT-based emotion classifier to evaluate the emotional sentiment of online responses. Our analysis of the volume shows a moderate relationship between real-world violence and Reddit activity, with a 15-day delay suggesting that responses are not instant, but rather delayed. Topic modeling shows that Reddit users tend to focus more on military actions and humanitarian crises following major high-fatality events, while identity-related discussion remains stable and protest-oriented themes receive limited attention. Sentiment analysis indicates that negative emotions, especially anger, dominate the conversation. However, these expressions remain relatively stable over time and are only weakly correlated with daily conflict severity. Together, these findings emphasise Reddit's selective and delayed response to conflict, with engagement being shaped more by emotionally striking or widely reported incidents than by the entire spectrum of violence.

Keywords: *Israel–Palestine conflict; Reddit discourse; Gaza; ACLED data; Topic modeling; Emotion modeling; Online response to violence.*

1 Introduction

1.1 Background

The Spring 2025 ongoing conflict in Palestine marks approximately 511 days after the Hamas-led October 7, 2023 attack on Israel. It represents a continued phase of an geopolitical conflict characterized by territorial asymmetry, recurring military operations, and chronic humanitarian distress. Gaza (Palestine), governed by Hamas since 2007 and subject to a blockade enforced by Israel and Egypt, has repeatedly experienced infrastructure collapse and civilian casualties due to military escalations (United Nations Office for the Coordination of Humanitarian Affairs, 2023; Roy, 2016; B'Tselem – The Israeli Information Center for Human Rights in the Occupied Territories, 2022). These flare-ups are fueled by a complex mix of political fragmentation, failed ceasefires, regional rivalries, and enduring ideological divisions (Milton-Edwards, 2008).

As it is a conflict in the 21st century, physical battlegrounds extended into digital environments. Access to traditional journalism became increasingly restricted and political narratives diverged (Borders, 2023), platforms like Reddit serves as alternative spaces for public interpretation, emotional expression, and real-time debate. These decentralized discussions are shaped not only by the visibility of events but also by platform logics, community norms, and algorithmic amplification (Weninger et al., 2013; Wolfsfeld, 2021). The overlap of real-world violence and online engagement raises critical questions: To what extent do social media narratives reflect, misrepresent, or distract from verified conflicts.

*Code available at https://github.com/tercasaskova311/Israel-Palestine-CSS-project

1.2 Focus

This research investigates the relationship between digital public discourse and verified conflict realities during the spring 2025 ongoing conflict in Palestine. We focus on comparing Reddit activity—measured through post and comments volume, dominant topics, and emotional tone—with conflict event data from the Armed Conflict Location and Event Data Project (ACLED). By aligning these two data sources, we aim to assess whether Reddit discussions reflect, distort, or diverge from the trajectory of real-world violence during the analyzed conflict period.

1.3 Relevance

The growing influence of social media on public perceptions of global conflicts offers both opportunities and risks. Platforms such as Reddit serve not only as forums for discussion but also as arenas where attention is distributed, emotions are intensified, and different interpretations change fast (Wolfsfeld, 2021). In conflict-affected regions where information is often unavailable or unreliable, online discussions can significantly influence international awareness, media coverage and humanitarian responses (Starbird, 2017). However, these online conversations are not neutral. They are influenced by the dynamics of the platform, emotional engagement and ideological perspectives. Analyzing how Reddit posts and comments address verified violence in Palestine can provide insight into broader societal dynamics, including selective empathy, misinformation, and the politicization of social media.

1.4 Time Frame

This study focuses on the period from 1 March to 20 June 2025, during which there was a sharp escalation in the Palestinian conflict. According to ongoing humanitarian reporting, this period was marked by large-scale airstrikes, attacks on aid convoys, and shelling of civilian infrastructure, resulting in significant casualties and displacement (United Nations Office for the Coordination of Humanitarian Affairs, 2025; Human Rights Watch, 2025). The scale and frequency of violence during these months led to widespread attention in both international media and digital platforms (Al Jazeera, 2025). This time frame allows us to analyze how peaks in real-world violence correlate with surges in online activity, offering a valuable opportunity to study lag effects, discourse dynamics, and emotional reactivity in digital spaces.

2 Literature Review

The focus of this study is to explore the interaction between Reddit discourse and verified conflict events during the 2025 Palestine conflict. We reviewed the literature on studies that use computational methods to analyze political polarization, conflict event data, or digital misinformation on online platforms. This was to ground our methodology and theoretical framing.

Building on the shift from the analysis of structured datasets to that of online discourse, Weber et al. (2013) examined ideological polarization on Twitter during the Egyptian revolution of 2011–2012. Using retweet graphs and community detection via the Louvain algorithm, they traced the growing divide between secular and Islamist users. Their study demonstrated the usefulness of social media data in capturing group dynamics and shifts in political discourse over time. Although our analysis does not utilize network graphs, we share their interest in discursive polarization, applying Latent Dirichlet Allocation (LDA) to Reddit posts to reveal thematic patterns in the discussion of conflict. However, unlike their approach, we link these themes directly to verified ground events, providing a dual-layered view of online narrative and offline reality.

In an effort to expand the computational social science analytical toolkit, Bail et al. (2018) introduced experimental methods involving the exposure of US Twitter users to opposing political views, followed by the measurement of changes in ideology and emotion.

By combining randomized controlled trials with sentiment analysis via LIWC (Linguistic Inquiry and Word Count), they discovered that exposure to opposing views frequently intensified political polarization. Although our study is observational rather than experimental, we explore a similar concept: how online emotion fluctuates in response to external stimuli, in our case real-world conflict events. Our approach goes beyond LIWC, using a BERT-based sentiment model fine-tuned for social media language to enable finer-grained emotion detection across a large Reddit corpus. While Bail et al. (2018) focused on affective shifts in US political contexts, our analysis targets emotional responses to violence, displacement, and humanitarian crises in a war zone.

The work of Guerra and Karakuş (2023), which analysed Reddit discourse during the Russian invasion of Ukraine, is most directly related to our research. Combining LDA topic modelling with lexicon- and transformer-based emotion classifiers, they demonstrated that spikes in fear and anger aligned with key military escalations. Their work demonstrated that Reddit can act as a channel through which public sentiment is expressed in response to armed conflict. Our study adopts a similar modelling pipeline, using LDA and a BERT-based emotion classifier. However, we build upon their approach by introducing a third analytical dimension: engagement volume. Using time-series and lagged correlation analysis, we examine how comment volume, topic prevalence, and emotional tone change before and after verified conflict events. This offers a more comprehensive view of the digital response to war.

In summary, the existing literature reflects a steady evolution from structured political event data to network-based online discourse analysis and causal studies of emotional exposure, culminating in Reddit-based thematic and emotional monitoring during real-world conflict. Our study synthesizes these approaches by combining three modalities, volume, topic, and sentiment, and aligning them with ACLED conflict timelines. This integrated temporal approach fills a significant gap in the literature by providing a multimodal event-aligned analysis of digital discourse in high-intensity conflict settings.

Table 1: Summary of Key Literature Reviewed

Source	**Research Design**	**Methods Used**	**Strengths**	**Limitations**
Raleigh et al. (2010)	Observational dataset of conflict events from news reports	Event coding (ACLED); extraction by time, location, actor, event type	High-resolution spatiotemporal data; widely adopted in policy and research	Excludes public discourse; limited to formal media content
Weber et al. (2013)	Observational study of Twitter during the Egyptian revolution	Retweet networks; Louvain clustering; hashtag tracking	Captures ideological clustering and network evolution	No link to real-world events; lacks sentiment or topic analysis
Bail et al. (2018)	RCT on exposure to opposing views	LIWC analysis; bot-driven tweet exposure; pre/post surveys	Causal inference on online discourse; controlled experimental design	U.S.-focused; not conflict-related or Reddit-relevant
Guerra and Karakuş (2023)	Observational Reddit study during Ukraine war	LDA topics; BERT- and lexicon-based emotion classification	Multi-modal discourse/emotion analysis; time-aligned with events	No volume or lag analysis; limited emotion labels

2.1 Research Question

Main Research Question:

This study focus on how Reddit discourse reflects or diverges from verified real-world conflict events in Palestine between March and June 2025. We analyzed this problematic through post volume, topic modeling, and sentiment analysis. Framed as a mixed-method study, the central question asks:

To what extent does digital engagement on Reddit temporally align with key moments of offline violence, and how do thematic and emotional patterns evolve in response to these events?

The **aim** is to explore whether observable spikes in attention and emotional expression on Reddit coincide with high-intensity real-world events, such as civilian casualties or humanitarian disruptions. We **expect partial alignment:** while post volume may surge around critical incidents, sentiment may remain more stable or be shaped by longer narrative arcs, platform dynamics, or ideological echo chambers. This would suggest a measurable gap between digital affect and lived conflict realities—a pattern with implications for public perception, media framing, and the role of social platforms during crises.

3 Project Design

This project uses a multi-method computational approach to compare Reddit discourse with verified conflict events during the 2025 ongoing conflict in Palestine. The central objective is to evaluate how Reddit engagement — measured in terms of volume, topic focus, and sentiment — corresponds with real-world violence. To achieve this, we integrate two datasets: Reddit posts -comments and ACLED events records.

The methodological choices we make are informed by a combination of theoretical and practical considerations. Reddit has been chosen because of its decentralised, long-form, and thematically organised discourse structure, which makes it suitable for extracting both narratives that occur organically and emotional responses. ACLED provides reliable ground truth for conflict events, enabling empirical validation. We use LDA to reveal latent topic structures in public discussions and transformer-based sentiment modelling to enable granular emotional classification beyond simple positive or negative polarity. Finally, time-series correlation and lag analysis are employed. These are used to examine the dynamic relationship between discussion and events over time. This approach strikes a balance between interpretability and computational precision, allowing for a comprehensive comparison between online narratives and offline realities.

3.1 Data Collection Strategy

This study analyzes two data sources to study the relationship between conflict events and online public discussion on the Israel–Palestine conflict:

- **Reddit Posts (Kaggle)** — representing online public discussion
- **ACLED Conflict Events** — representing real-world violent events

3.2 Methodology

Preprocessing

Reddit Dataset The Kaggle Reddit dataset[1], covers posts from **March 1 to June 20, 2025** and preprocessed using a multi-stage pipeline to prepare high-quality input for sentiment analysis and topic modeling. For each entry, the comment text, post title, and body were merged into a single field and truncated to a maximum of 500 words to standardize input length and manage resources.

[1] `https://www.kaggle.com/datasets/asaniczka/reddit-on-israel-palestine-daily-update`

The Preprocessing involved two main stages. First, regular expressions were used to lowercase text, remove URLs, punctuation, numbers, and excess whitespace. Next, spaCy was employed for part-of-speech tagging, lemmatization, and removal of stopwords and short tokens (fewer than three characters), while preserving sentence structure. All steps were executed in batch mode for efficiency. The cleaned text was stored alongside the original data to ensure consistent input for all subsequent analyses.

ACLED Dataset ACLED conflict data[2] were filtered for events in both Gaza and the West Bank over the same period of time. Including columns:

- `event_date`, `event_type`, `sub_event_type`
- `actor1`, `civilian_targeting`, `location`
- `notes`, `fatalities`, `population_best`

Metadata and irrelevant fields were excluded. A new variable, `event_category`, was created based on `event_type` and `sub_event_type` to group events into two types:

- **Combat** — including battles and explosions/remote violence
- **Civilian Harm** — including direct attacks on civilians

Volume Comparison

Volume comparison refers to analysis, where we compare how much activity occurs in two distinct data sources over time. In our case, the number of conflict events in Palestine (based on verified records from Non-Profit organization ACLED) and the volume of Reddit comments discussing the Israel-Palestine conflict. By aligning these data sources, we can analyze whether online attention increases when real-world violence spikes. This approach helps us understand how public engagement unfolds. We analyze if it is immediate, delayed, or selective in what it reacts to.

In order to investigate the relationship between off-line events and online discussion, we compared the daily volume of conflict events reported by ACLED and Reddit comment activity. These time series were first aggregated at the daily level. To reduce short-term fluctuations and highlight longer-term patterns, we applied a 7-day centered rolling average to each series. The smoothed values were then normalized to a common scale, enabling direct visual comparison despite the differing units and magnitudes of the two data sources.

Through this comparison, it can be determined whether spikes in conflict activity are reflected in increased Reddit engagement and the extent to which these patterns have been aligned in time. Identifying the timing and nature of such relationships directly contributes to answering our research question, which is to what extent Reddit discourse mirrors or diverges from real-world conflict dynamics.

Reddit Topic Modeling with LDA

Latent Dirichlet Allocation (LDA) is a probabilistic topic modeling technique used to uncover abstract topics in large texts. It assumes that every Reddit post is a mixture of multiple topics, with each topic defined by a set of words. Instead of manually reading thousands of posts, LDA helps us to identify the main themes in the discourse, such as military actions, humanitarian concerns or political narratives, and to track how these themes evolve over time.

To analyze the topic structure of Reddit discussion surrounding the Israel–Palestine conflict, we applied LDA to a cleaned dataset of Reddit posts. Text was converted into a document-term matrix using `CountVectorizer`, configured with `min_df = 10` and

[2] `https://acleddata.com/`

`max_df = 0.9` to exclude rare or common terms. English stop words were also removed. The number of topics was set at five. This was chosen based on the observation that, beyond this point, coherence scores plateaued, indicating an absence of improvement in model quality.

Each topic is represented by a probability distribution of words, and each Reddit post is assigned a probability distribution across the five topics. To improve interpretability, we summarise each topic using its most probable keywords. **Top 5 LDA Topics and Representative Keywords:**

- **Topic 3:** israel, iran, trump, israeli, attack, netanyahu, country, gaza, hamas, nuclear
- **Topic 0:** israel, jews, people, palestinians, state, jewish, palestine, land, palestinian, arab
- **Topic 4:** gaza, people, israel, hamas, aid, israeli, say, war, like, food
- **Topic 1:** israel, hamas, civilian, people, gaza, israeli, war, kill, idf, say
- **Topic 2:** israel, people, war, like, want, gaza, think, protest, world, genocide

Instead of assigning each post to a single dominant topic, we aggregated the full topic probability vectors for all posts on each day. This produced continuous daily estimates of the relative prevalence of each topic in the discourse. These time series were then smoothed using a seven-day rolling average centred on the middle day to reduce short-term fluctuations and highlight longer-term trends.

This method allows us to evaluate how specific Reddit discussion topics align, or fail to align, with the intensity and nature of real-world conflict events, thus contributing directly to our central research question.

Sentiment Analysis

To complement our topic modeling approach, we conducted an emotional tone analysis of Reddit discussions related to the conflict using computational sentiment analysis. Our objective was to examine the relationship between real-world conflict events, as documented by the ACLED Palestine Monitor, and the emotional tenor of public discourse on Reddit. To achieve this, we developed a multi-step analytical pipeline integrating natural language processing, time series analysis, and statistical inference.

We began by pre-processing and aggregating two core datasets: Reddit comments discussing the conflict and event data from the ACLED Palestine Monitor. The Reddit data was subjected to standard text-cleaning procedures and subsequently analyzed using a pre-trained BERT-based emotion classification model, fine-tuned for social media contexts. Specifically, we employed the model developed by Hartmann (2022), available via Hugging Face[3]. This transformer-based model was chosen over traditional lexicon-based tools such as VADER due to its enhanced performance on informal, context-rich social media text, such as that found on Reddit.

Unlike VADER, which relies on predefined sentiment lexicons and is best suited for short, neutral, or formal text, the BERT-based model leverages deep contextual embeddings to better capture slang, sarcasm, and emotionally charged language. Trained on diverse datasets, including Reddit and Twitter, it is particularly well-equipped to analyze the nuanced emotional content of user-generated discussions. Importantly, this model outputs probabilistic scores for seven discrete emotions, allowing us to move beyond simple sentiment polarity (positive/negative) and toward a more granular understanding of emotional expression. This makes it highly effective for tracking changes in public sentiment in response to real-world events (Hartmann, 2022).

Emotion scores were aggregated at the daily level, generating a time series of average emotion intensities. Similarly, the ACLED dataset was processed to yield daily counts of conflict events, fatalities, and event types. The two datasets were then merged on the

[3] `https://huggingface.co/bhadresh-savani/bert-base-uncased-emotion`

temporal dimension, enabling direct comparison between daily emotional patterns and conflict dynamics.

This merged dataset served as the foundation for a series of quantitative analyses. We first conducted correlation analyses to evaluate the strength and direction of associations between conflict metrics (e.g., fatalities, event counts) and emotional responses (e.g., anger, fear, sadness). To investigate temporal dynamics, we applied lagged correlation techniques and event study designs, assessing whether emotional reactions precede or follow significant conflict events.

Further, we used statistical testing to compare emotional responses on high-fatality days versus typical days, and examined how emotional distributions vary by ACLED event type. Visualization techniques—including time series plots, scatter plots, and heatmaps—were employed to present key findings and aid in the interpretation of complex patterns.

4 Results

This section presents the main empirical findings from the three analytical approaches introduced in Section 3. Specifically, we report results from (1) a comparison between Reddit comment volume and conflict intensity, (2) topic modeling using Latent Dirichlet Allocation (LDA), and (3) sentiment analysis using a transformer-based emotion classifier. Each subsection discusses the patterns observed and their relevance to understanding public discourse around the Israel-Palestine conflict.

Volume Comparison

Between March 1 and June 20, 2025, the average number of ACLED-reported conflict events per day was 102.79 (maximum = 144), while Reddit comment volume averaged 3,681.05 per day (maximum = 10,098).

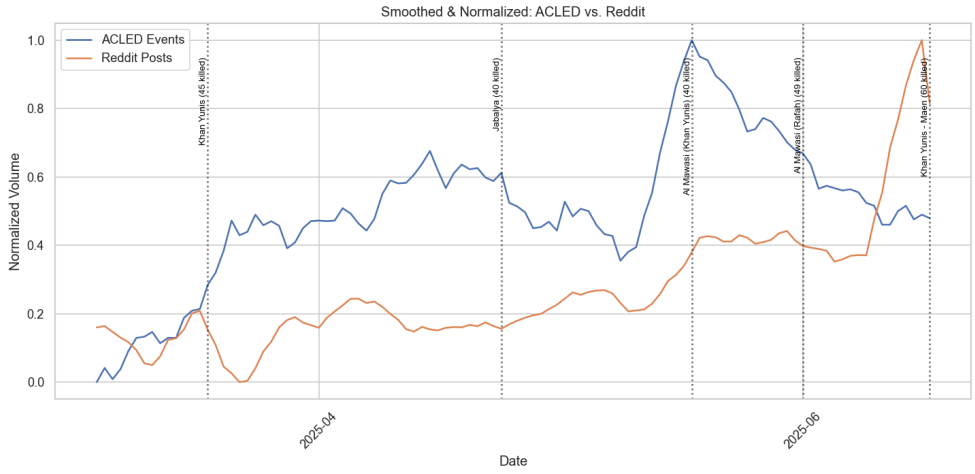

Figure 1: Normalized and smoothed time series comparison between Reddit comment activity and ACLED-reported conflict events.

The Pearson correlation between the smoothed and normalized time series was $r = 0.37$ ($p < 0.001$), indicating a moderate positive relationship between conflict intensity and Reddit engagement. To explore possible lagged effects, we conducted a cross-correlation analysis using z-score standardized data (Figure 2). The strongest correlation was found at a lag of 15 days ($r = 0.56$), suggesting that Reddit activity tends to increase approximately two weeks after spikes in conflict events. A smaller, yet statistically significant, correlation at lag 0 ($r = 0.39$) indicates some degree of immediate responsiveness.

Although Reddit and ACLED data originate from fundamentally different domains (digital discourse and physical conflict), there is a measurable relationship between the

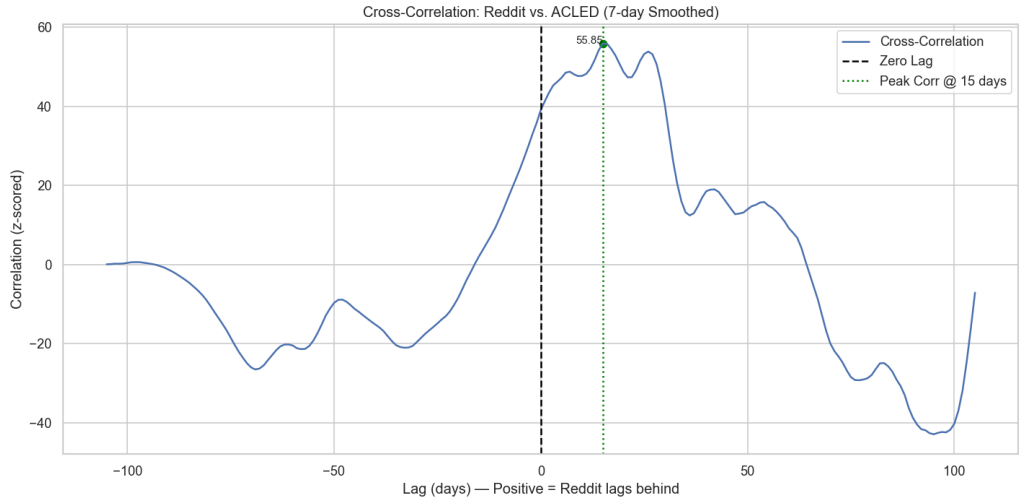

Figure 2: Cross-correlation between ACLED conflict events and Reddit comment activity.

two. The observed 15-day delay aligns with broader media amplification cycles. This may indicate that Reddit functions more as a space for reflective analysis than for immediate reaction.

Comment volume also spiked on specific dates corresponding to widely reported incidents involving high civilian casualties. These high-engagement moments, verified through fatalities and reported in global media, are summarized below:

- **June 17, 2025** – *10,719 comments*: Israeli tanks fired on a crowd awaiting aid trucks in Khan Younis, killing at least 59 Palestinians and wounding over 200 civilians (Reuters, 2025b).
- **June 1, 2025** – *5,393 comments*: Israeli forces opened fire near an aid distribution site in Rafah, resulting in at least 31 deaths and approximately 170 injuries.
- **May 18, 2025** – *5,000 comments*: A large-scale Israeli airstrike campaign killed over 150 Palestinians in a single day.
- **April 24, 2025** – *3,202 comments*: Airstrikes in Jabalia refugee camp killed 44 Palestinians and damaged a children's hospital (Reuters, 2025a).
- **March 18, 2025** – *4,236 comments*: A sudden escalation broke a two-month truce, with over 400 fatalities reported, including 263 women and children.

These spikes suggest that Reddit users are most responsive to:

1. **Aid-related civilian harm** (e.g., June 17, June 1), reflecting acute public concern over humanitarian violations.
2. **Major military escalations** (e.g., March 18, May 18), which appear to drive substantial attention and discourse.
3. **Sustained violence** (e.g., April 24), even when not marked by a singular catastrophic event.

Overall, these findings support the idea that Reddit discourse responds to real-world conflict. This is particularly the case when fatalities are high or civilians are visibly targeted. However, the absence of spikes on certain days of high conflict suggests that engagement is also influenced by other factors, such as media coverage, political framing or perceived emotional significance. The time lag and inconsistency imply that Reddit discussions do not offer a direct reflection of real-time violence, but rather a filtered, sometimes delayed reflection, shaped by broader digital attention dynamics.

LDA Topic Trends

Figure 3 presents the smoothed daily prevalence of each LDA topic, plotted alongside ACLED-reported fatalities. These trends help evaluate whether Reddit discourse follows, precedes, or diverges from real-world conflict severity.

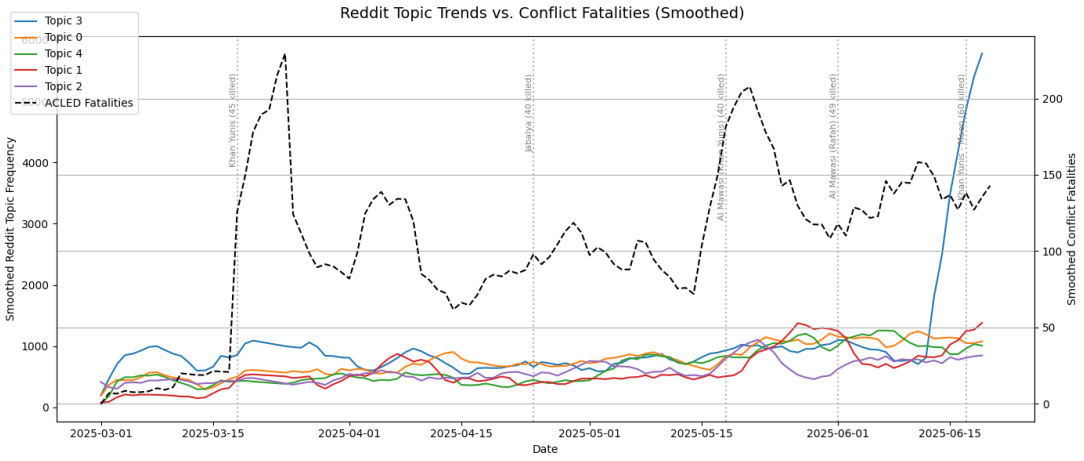

Figure 3: Temporal trends of the five dominant Reddit topics extracted using LDA.
Note: The proportions of the Reddit topic are smoothed using a seven-day rolling average and plotted alongside the number of daily fatalities reported by ACLED in Gaza and the West Bank. Fatality counts were used instead of total event counts to better reflect conflict severity.

Interpretation of LDA Topic Trends:

- **Topic 3 (Geopolitical discourse):** *israel, iran, trump, israeli, attack, netanyahu, country, gaza, hamas, nuclear* — This topic shows a pronounced spike following the June 17 Khan Younis aid strike, suggesting a shift in Reddit discourse toward geopolitical blame, state actors, and broader international narratives.

- **Topics 1 (Military conflict)** and **4 (Humanitarian crisis):** *israel, hamas, civilian, gaza, war, kill, aid, food, people* — These show increases from mid-May through early June, corresponding with repeated high-casualty events in Rafah and Al Mawasi. This indicates strong reactivity to both direct violence and humanitarian issues.

- **Topic 0 (Identity discourse):** *israel, jews, palestinians, jewish, arab, land* — This topic remained relatively stable throughout the observed period. It appears to remains consistent despite daily events.

- **Topic 2 (Global perception and protest):** *protest, world, genocide, think, want* — This is the least common and reactive topic. This probably reflects broader conceptual frameworks rather than immediate, emotionally driven responses to violence.

In summary, Reddit discussions show clear shifts in topic following major conflict events, especially towards military violence and humanitarian crises. The ongoing discussions about identity suggest an underlying ideological framework shaping the conversation, while lower engagement in protest-related topics may be due to Reddit's audience or its lack of immediate activism capabilities.

These results address the research question by showing how the content of Reddit discussions is shaped by both events and underlying structures. While some topics align closely with spikes in fatalities, others remain unaffected, highlighting the selective amplification of certain narratives over others. More sophisticated modelling is required for causal inference, in order to move beyond the current descriptive approach.

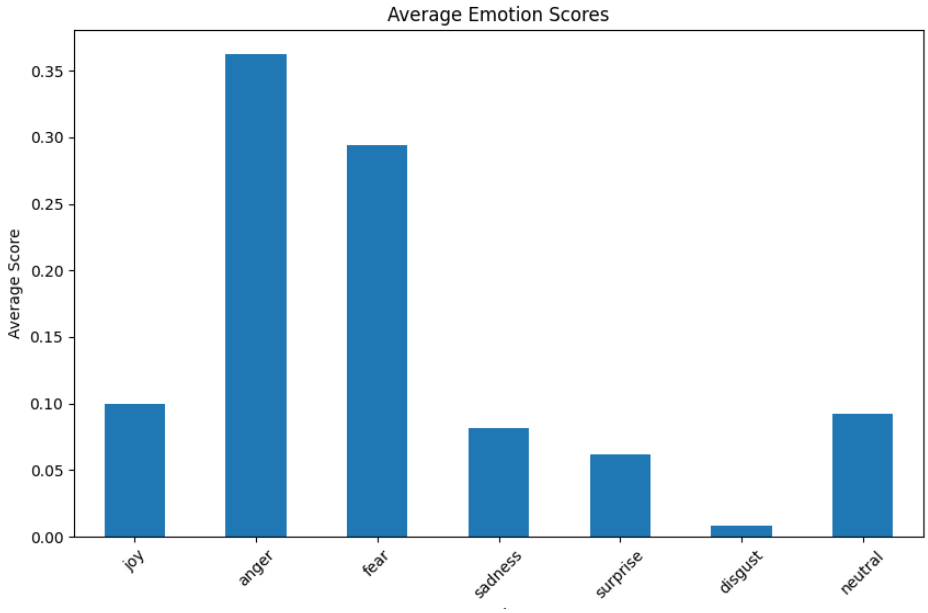

Figure 4: Mean Emotion Scores in Reddit Discussions of the Israel-Palestine Conflict. Probabilistic emotion scores (range: 0–1) were computed using BERT-based sentiment analysis of all Reddit comments (N=417,567) without temporal aggregation.

Sentiment Analysis

The sentiment analysis pipeline processes each Reddit comment through the BERT model to generate probabilistic scores for each emotion category. These scores are then aggregated at the daily level, enabling temporal alignment with conflict event data from the ACLED Monitor. As illustrated in Figure 4, the resulting distribution of emotion reveals a pronounced dominance of negative affect in discourse surrounding the Israel-Palestine conflict. These findings underscore the overwhelmingly negative tone of Reddit discussions, which aligns with the polarizing and traumatic nature of the conflict.

We examined the relationship between emotional responses and conflict intensity by correlating daily fatalities from the ACLED dataset with aggregated Reddit emotion scores, considering both same-day and one-day lagged effects. As shown in Table 2, Same-day correlations were generally weak, with the strongest negative associations observed for neutral and disgust. The positive correlations for sadness, anger and fear were minimal. When examining lagged correlations, testing whether emotional responses follow conflict events with a one-day delay, the results remain modest, with only fear (r = 0.117) and anger (r = -0.100) showing weak significance. Overall, the results suggest that while some emotions are modestly associated with conflict intensity, the emotional tone of Reddit discourse is only weakly responsive to daily variations in violence, indicating that other factors likely influence online emotional expression.

The timeline analysis (Figure 5) shows that while Reddit comment volume rises notably during periods of heightened conflict—particularly in mid-March, mid-May, and June, emotional responses exhibit only modest variation. Anger consistently dominates, with occasional spikes in fear aligning with high-fatality events, suggesting partial emotional sensitivity to violence. Joy shows mild increases following days with fewer civilian-targeted attacks, indicating a possible response to de-escalation. Overall, Reddit engagement appears more reactive to conflict intensity than emotional tone, which remains relatively stable, hinting that broader contextual factors—such as media coverage or discourse dynamics, may play a larger role in shaping the public sentiment.

To further assess whether emotional expressions differed meaningfully on days of ex-

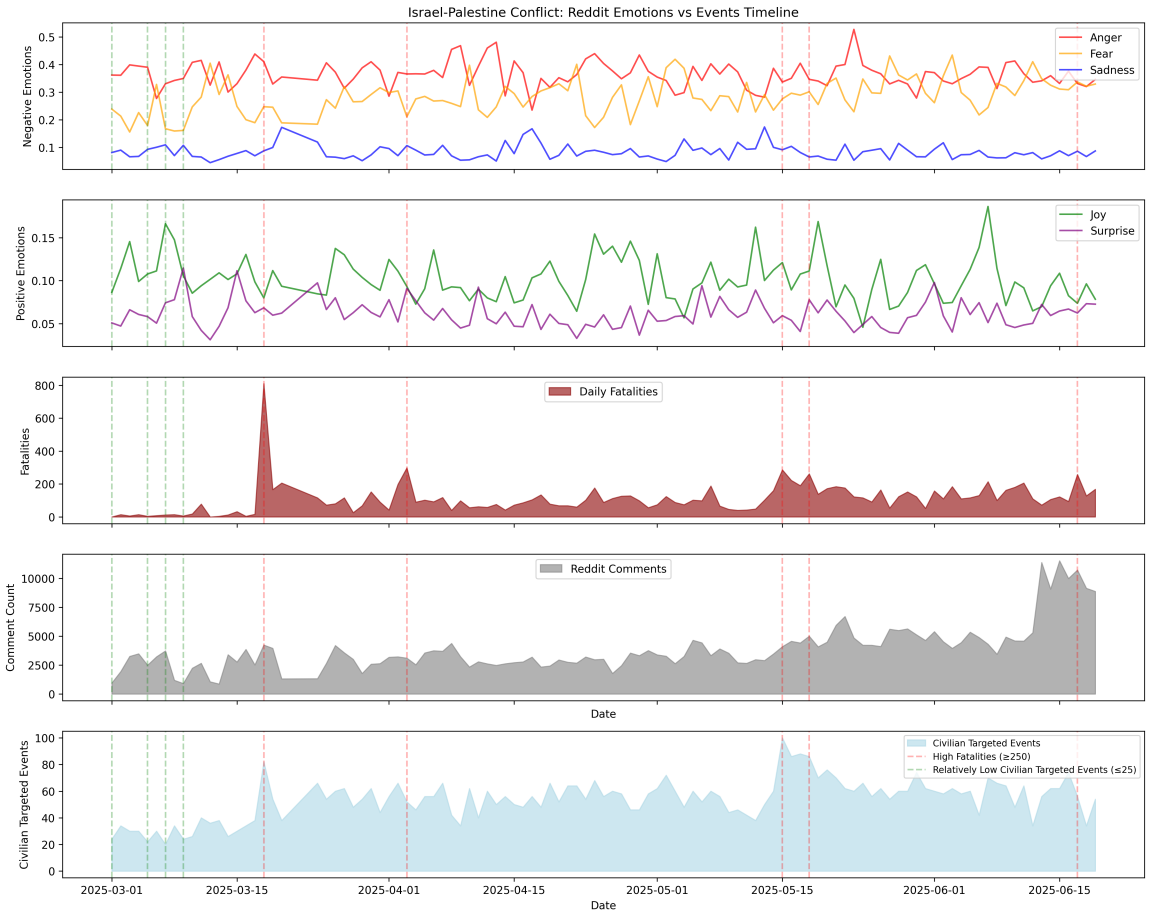

Figure 5: Daily average Reddit emotion scores with conflict indicators. Red dotted lines mark high-fatality days (≥ 250 fatalities); green lines mark days with low civilian targeting (≤ 25 events). While comment volume increases during major conflict periods, emotional expressions, especially anger, remain relatively stable, with only slight shifts in fear and joy.

Table 2: Same-day and one-day lagged Pearson correlation coefficients (r) between emotion scores and daily fatalities.

Emotion	Same-day r	Lagged r (t–1)
Neutral	**–0.234**	–0.064
Disgust	–0.139	–0.025
Sadness	0.093	0.095
Anger	0.087	**–0.100**
Fear	0.048	**0.117**
Joy	–0.070	–0.036
Surprise	–0.021	–0.028

treme conflict, we conducted a Mann-Whitney U test comparing emotion scores between high-fatality days (defined as the top 10% of days by fatalities, i.e., ≥ 188.6 deaths) and the remainder of the observation days. This non-parametric test revealed no statistically significant differences across any emotion category, including anger (p = 0.76), fear (p = 0.79), joy (p = 0.91), and others (all $p > 0.3$). These results support the earlier observation that, while Reddit engagement intensifies during periods of violence, the distribution of emotional responses remains relatively stable. This suggests that emotional tone on Reddit may be shaped more by broader discourse patterns than by short-term fluctuations in conflict intensity.

Unlike Guerra and Karakuş (2023), who observed clear sentiment spikes aligned with conflict events, our statistical analysis indicates that emotional expressions on Reddit do not significantly vary on high-fatality days. While Reddit engagement intensifies during violent incidents, the emotional intensity—particularly fear and anger—remains persistently high and stable, suggesting that overall sentiment is shaped more by sustained and broader discourse patterns than by immediate conflict fluctuations. This interpretation aligns with Huh and Park (2024)'s analysis of social media sentiment during the 2016 U.S. election, which found that emotional responses are more strongly influenced by ongoing partisan discourse and media framing than by short-term real-world events.

Limitations

While the results provide empirical insights into the relationship between Reddit discourse and real-world conflict events, several limitations should be noted. First, the dataset only includes public Reddit comments and may exclude deleted, edited, or shadow-banned content, introducing sampling bias. Second, the BERT-based sentiment model, while robust, may struggle with detecting sarcasm, mixed emotions, or culturally specific expressions common in Reddit discourse. Finally, the correlational nature of the analysis precludes causal inference—emotional responses may be shaped by external media narratives, not just direct exposure to conflict events.

5 Conclusions Based on Empirical Analysis

5.1 Empirical Conclusions

In this study we analyzed the relationship between Reddit discussions and real-world conflict events in Palestine between March 1 and June 20, 2025. The analysis focused on three key areas: the volume of comments versus the frequency of events; topic content; and sentiment.

Volume. Reddit comment activity followed the frequency of ACLED-reported conflict events. This relationship was moderately strong. A cross-correlation analysis revealed a delay of approximately 15 days, with online engagement peaking around two weeks after

spikes in violence. It is suggested by this that Reddit is responsive but not immediate, likely due to different reasons, for example media amplification cycles or the platform's slower-paced discussion style.

Topics. Topic modeling (LDA) findings revealed that Reddit discussions changed in response to major conflict events:

- *Military and humanitarian topics* increased after days of high mortality.
- *Geopolitical topics* surged following politically sensitive incidents.
- *Identity-related themes* remained relatively unchanged.
- *Protest-related discourse* stayed low and did not clearly react to specific events.

Sentiments. The analysis reveals that Reddit discourse surrounding the Israel–Palestine conflict is dominated by negative emotional expressions, with anger emerging as the most consistently prevalent sentiment. Despite significant spikes in comment volume during periods of heightened conflict, the distribution of emotional categories remains relatively stable over time. Correlational analysis indicates only weak associations between daily conflict fatalities and emotional responses, with the strongest lagged effects observed for fear (r = 0.117) and anger (r = –0.100), though these remain modest in magnitude. Furthermore, non-parametric testing (Mann-Whitney U) finds no statistically significant differences in emotional expression between high-fatality days and the remainder of the observation period, reinforcing the conclusion that the emotional tone of Reddit discussions does not shift dramatically in response to acute changes in conflict intensity.

5.2 Connection to Research Question

The central research question in this project was:

> *To what extent does Reddit discourse — measured through post volume, dominant topics, and sentiment—align with or diverge from verified real-world conflict events in Palestine?*

The findings suggested that Reddit discussions about the Israel–Palestine conflict generally align with real-world events, but not in immediate way. While the platform tends to respond most noticeably to severe or highly visible incidents—especially those involving civilian casualties—the overall emotional tone remains relatively stable, with only modest shifts in response to daily fluctuations in conflict intensity. This suggests that Reddit reflects only part of the conflict narrative, emphasizing emotionally charged or widely reported events while overlooking others.

5.3 Evidence-Based Support

The conclusions are drawn from the following analytical results:

- **Time series alignment** between Reddit activity and ACLED conflict records demonstrates overlapping patterns, especially during periods of intensified conflict.
- **Cross-correlation analysis** reveal modest yet detectable lag effects, implying that online discussions sometimes trail real-world events by short temporal intervals.
- **Topic modeling trends** show shifts in Reddit discussion topics over time, including spikes in politically charged topics during conflict peaks.
- **Sentiment patterns** indicates that emotional tone on Reddit remains relatively stable, with no consistent increase in emotional polarity during periods of intense violence.

- **Correlation analysis (Table 2)** demonstrates low and statistically insignificant correlations between conflict fatalities and sentiment measures, for both same-day and lagged comparisons.
- **Mann-Whitney U test** results (p-values > 0.3 across all emotion categories) show no statistically significant change in sentiment distribution during conflict versus non-conflict periods.

These findings indicate that while Reddit discussion volume may increase during conflict events, its thematic and emotional content does not consistently reflect the severity or timing of real-world violence. The platform shows selective engagement, with limited emotional reactivity and weak correlation to conflict intensity.

References

Al Jazeera. (2025). *Live updates: Gaza conflict escalates march–june 2025.* Retrieved from https://www.aljazeera.com/news/longform/live-updates/gaza-2025 (Accessed June 2025)

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., ... Volfovsky, A. (2018, September). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. Retrieved from https://www.pnas.org/doi/10.1073/pnas.1804840115 doi: 10.1073/pnas.1804840115

Borders, R. W. (2023). *Palestinian territories: Journalists caught in the crossfire.* Retrieved from https://rsf.org/en/palestinian-territories (Accessed June 2025)

B'Tselem – The Israeli Information Center for Human Rights in the Occupied Territories. (2022). *Gaza strip.* Retrieved from https://www.btselem.org/gaza_strip (Accessed June 2025)

Guerra, A., and Karakuş, O. (2023). Sentiment analysis for measuring hope and fear from reddit posts during the 2022 russo-ukrainian conflict. *Frontiers in Artificial Intelligence*, *6*, 1163577. Retrieved from https://www.frontiersin.org/articles/10.3389/frai.2023.1163577/full doi: 10.3389/frai.2023.1163577

Hartmann, J. (2022). *Emotion english distilroberta-base.* https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/.

Huh, C. U., and Park, H. W. (2024). Setting the public sentiment: Examining the relationship between social media and news sentiments. *Systems*, *12*(3). Retrieved from https://www.mdpi.com/2079-8954/12/3/105 doi: 10.3390/systems12030105

Human Rights Watch. (2025). *Escalation in gaza: Civilian harm and humanitarian access.* Retrieved from https://www.hrw.org/news/2025/04/30/gaza-civilian-harm (Accessed June 2025)

Milton-Edwards, B. (2008). *The israeli-palestinian conflict: A people's war*. Routledge.

Raleigh, C., Linke, A., Hegre, H., and Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset. , *47*(5), 651–660. Retrieved 2025-06-30, from http://www.jstor.org/stable/20798933 (Publisher: Sage Publications, Ltd.)

Reuters. (2025a). *Israeli fire kills at least 44 people in gaza, hits a police station.* Retrieved from https://www.reuters.com/world/middle-east/israeli-fire-kills-least-26-people-gaza-hits-police-station-2025-04-24/ (Accessed June 2025)

Reuters. (2025b). *Israeli tank shelling kills dozens awaiting aid in gaza.* Retrieved from https://www.reuters.com/world/middle-east/israeli-tank-shelling-kills-45-people-awaiting-aid-trucks-gaza-ministry-says-2025-06-17/ (Accessed June 2025)

Roy, S. (2016). *The gaza strip: The political economy of de-development*. Institute for Palestine Studies.

Starbird, K. (2017). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 230–239.

United Nations Office for the Coordination of Humanitarian Affairs. (2023). *Gaza strip: Humanitarian snapshot.* (https://www.ochaopt.org)

United Nations Office for the Coordination of Humanitarian Affairs. (2025). *Gaza situation report – march to june 2025.* (https://www.ochaopt.org/reports)

Weber, I., Garimella, V. R. K., and Batayneh, O. (2013). Secular vs. islamist polarization on twitter in egypt. In *Proceedings of the 2013 ieee/acm international conference on advances in social networks analysis and mining (asonam)* (pp. 290–297). IEEE. doi: 10.1145/2492517.2492567

Weninger, T., Zhu, X., and Han, J. (2013). An exploration of discussion threads in social news sites: A case study of the reddit community. *Advances in Social Networks Analysis and Mining (ASONAM)*, 579–583.

Wolfsfeld, G. (2021). *Making sense of media and politics: Five principles in political communication*. Routledge.