# LexiconLadies at FIGNEWS 2024 Shared Task: Identifying Keywords for Bias Annotation Guidelines of Facebook News Headlines on the Israel-Palestine 2023 War

**Yousra Alaa El-Ghawi, Abeer Ehab Marzouk, Aya Mohamed Khamis**
Alexandria University, Egypt
yousraghawi, abeerehab1324, ayamohamedkhamisali@gmail.com

## Abstract

News bias is difficult for humans to identify, but even more so for machines. This is largely due to the lack of linguistically appropriate annotated datasets suitable for use by classifier algorithms. The FIGNEWS Subtask 1: Bias Annotation involved classifying bias through manually annotated 1800 headlines from social media. Our proposed guidelines investigated which combinations of keywords available for classification, across sentence and token levels, may be used to detect possible bias in a conflict where neutrality is highly undesirable. Much of the headlines' percentage required contextual knowledge of events to identify criteria that matched biased or targetted language. The final annotation guidelines paved the way for a theoretical system which uses keyword and hashtag significance to classify major instances of bias. Minor instances with bias undertones or clickbait may require advanced machine learning methods which learn context through scraping user engagements on social media.

## 1 Introduction

Headlines are the first and often the only element of news which readers interact with (Ecker et al., 2014), especially on social media. What this meant for online and social media journalism was that readers could now filter their 'feeds' or 'timelines'. Most news is not viewed by a general audience anymore and instead are pushed by algorithms further or closer to the viewer (Mohammadinodooshan and Carlsson, 2023), depending on their preferences and ideologies. From there we could explore how journalists and public figures used targeted language to cater to specific groups or parties (in the case of political news coverage) or to attempt to reach and influence the mindset of other groups. Such endeavor is referred to as strategic communication and is a contributor to media bias (Finkbeiner, 2024).

To track news bias, researchers often devise their own datasets. Many of the available datasets focus their efforts on bias detection on the document level, by news outlet source, or by using automated methods which lack proper inference of implicit bias. Examples of such methods, and perhaps the closest and most elaborative studies on keyword classification of news sentiment (leading to bias) are the works including Hamborg, Spinde, and Gipp 2021. In their often combined research efforts on news bias in conventional news (Spinde et al., 2021; Hamborg et al., 2021) on social media (Spinde et al., 2023), they bridged the context gap between news outlets and readers. One of their methods involved building a keyword lexicon to utilize in the training of word embeddings. Attempts to identify bias triggers discovered the possibility of relying on linguistic features for shallow or simple bias detection (Lim et al., 2018), providing some characteristics of bias-inducing words. Lim et al. (2020) crowdsourced again a dataset of news, this time consisting of 966 sentences for 4 different events, concluding that prior knowledge about the context surrounding news is a major factor in bias detection. Context is a difficult subject. It cannot be easily inferred from specific keyword presence only. Ultimately, it depends on the readers' ideologies, stances, and knowledge (Hube and Fetahu, 2018).

Research on the biased language of reporting on the Palestinian-Israeli conflict has been studied in several attempts, however, the resulting datasets are relatively small and may not be sourced from a variety of news sources (Deegan et al., 2028; Al-Agha Abu-Dahrooj, 2019) (Deegan et al., 2018; Al-Agha and Abu-Dahrooj, 2019). Better bias detection classification models require valid training and evaluation of diverse corpora of news on the conflict. This paper tries to explore the development of bias recognition specifics for this conflict, especially on social media.



*Proceedings of The Second Arabic Natural Language Processing Conference*, pages 561–566

The FIGNEWS Subtask 1: Bias Annotation was done through a linguistic viewpoint, in an attempt to compile a sound semantic and contextually backed understanding of the discourse differences surrounding the Israeli war on the Gaza Strip (Zaghouani et al., 2024). One important aspect of the dataset's headlines is that they are extracted from social media, namely Facebook. Social media news has the additional feature of permitting user interaction (when applicable) under news to contextualize reporting. Depending on the audience reaction, biased news may be balanced to objectivity through user input or they may be further amplified towards the biased direction they intended (Spinde et al., 2023).

Facebook as a platform has long been complicit in giving emphasis to Israeli and Zionist news and voices (Alhossary et al., 2023). The biased content moderation of Palestinian content has been studied during several Israeli campaigns on Palestinian territories, such as the Sheikh Garrah crisis (Elmimouni et al., 2024) and the Guardian of the Walls operation (Abushbak et al.). The extent to which social media sites seek to censor or "reduce the visibility" (Gillespie, 2022) of certain ideologies, including pro-Palestinian posts and headlines, has caused several commentators to similarly censor their intentions while maintaining bias towards the Palestinian cause.

Several headlines required external knowledge of the surrounding events and context to correctly label them as biased (Bonet-Jover et al., 2023). Therefore the task is pragmatic by nature, with semantic considerations to the underlying meanings of words and expressions, as well as media literacy and considerations of culture and audience (Ricketts, 2024). Most of the previous attempts at annotating datasets regarding news headlines focused on credibility and misinformation identification (Bountouridis et al., 2019). While the aspect of veracity is a part of bias detection, our focus is on the required knowledge of the language used to identify biased headlines. As language models do not fully acknowledge cultural context, and rather measure the statistics surrounding entities and events, we attempt to illustrate a theoretical system based on overall sentence sentiment and token-level signs that may be of use to build models of better bias classification.

We ask the following: if a system were to analyze the resulting short (headline) and long-form (posts) news, what combination of keywords would be required to infer the implicit bias or lack thereof? Furthermore, how much does world context, user engagement with social media news, and media literacy affect the overall detection of possible bias?

## 2 Methodology

### 2.1 Development of Annotation Guidelines

The process of discourse-pragmatic annotation guidelines creation usually goes through stages: pilot annotations of the dataset, the gathering of target words of interest which contribute to specific labeling (Gries and Berez, 2017). We developed guidelines based on two levels of meaning and relevance: the sentence-level (how biased is the overall sentence?), and for refinement, the token-level (which specific words may be used to validate the sentence-level labeling?). We also sought to implement our knowledge of the dataset's main news events as mentioned by the FIGNEWS task organizers. By combining keywords of sentiment markers, events, and targeted language against others, we could refine most of our annotations and maintain a fair level of consistency. To decide on a preliminary lexicon of keywords based on relevance towards or against the parties of the war, we first looked up any attempts to measure word frequencies within the Israel-Palestine conflicts (Sevón, 2020; Majzoub, 2021; Alashqar, 2024;). After labeling a number of pilot annotations, more repeating themes and words were discovered. The most effective words and their collocates would be represented in the discussion. Our guidelines document construction scored a total of 7 on a scale of 10 points, while the developed guidelines reached a score of 0.6732, which included the IAA Kappa score of 37.2 and the weighted document score.

### 2.2 Data Annotation Process

The data provided by the FIGNEWS Subtask 1 (Zaghouani et al., 2024), a subset of the full corpus entailing multilingual headlines and their advertisement posts from Facebook, equated to around 13,500 posts separated into 15 batches. We manually annotated two batches, amounting to 1800 posts. Another 200 headlines per annotator were separately annotated as pilot annotations, and to measure consistency amongst the annotators. The data collected covers major events of the conflict in 5 languages (Arabic, English, Hebrew, French, Hindi). Columns for the dataset are separated into Batch, Source Language, ID, Type (Main batch or

IAA batch), Text, English MT, Arabic MT, Annotator ID, and Bias (annotation assignment field). The keywords for the events are major in determining context that may be a supplement in deducing bias.

The main dataset annotations were done while considering the developed keyword-based guidelines. We set our examples as per table (1). Annotators would either simply iterate through posts or filter columns by keyword corresponding to a certain label. The second approach saved more time, but was more prone to skip headlines which lacked specific keywords.

### 2.3 Inter-Annotator Agreement (IAA) Analysis

To ensure the consistency of annotations within an ethical frame, at least two annotators were required to annotate at least 200 headlines and posts as pilot annotations. Following the annotation of 400 main dataset posts, as well as 400 combined posts within the Inter-Annotator Agreement (IAA) sheets, pilot annotations were ready. What proceeded was a meeting for discussion of certain inconsistent label statistics. The major inconsistencies discovered were in regards to the "Unbiased" and "Biased against others" labels. We decided for the guidelines to use the keyword significance method, free of classification models, as we thought of this task as the exploration of pragmatic human understanding to apply later to machine learning. After tallying of the results by FIGNEWS organizers, our IAA Kappa score within the annotation team was at 37.2, rated as "fair". Meanwhile, the F1 Bias IAA reached a score of 73.4.

## 3 Team Composition and Training

The annotation team responsible for the task consisted of 3 women (ages range from 22 to 24, "generation "Z"") and recent graduates in applied linguistics with a native Arabic background of different dialects. All members were proficient in English and two annotators had working knowledge of French. Personal preference was granted for the choice of language to annotate headlines in. Most of the annotations were done in the MT English version, therefore the final guidelines used English keywords to recognize and filter possible bias. Two annotators had very little annotation experience prior to the task. Further training included offline discussions on labeling specifics and deciding on where to scan headlines for potential bias inducing words.

## 4 Task Participation and Results

We tried to produce results of higher quality than quantity, in which we implemented our contextual and linguistic knowledge to the labeling efforts. This knowledge included: 1. Known social media engagement tactics, 2. Known political and humanitarian stances of news outlets and countries, and 3. One's own bias regarding the war.

The highest recorded label was "Biased against Palestine", with a percentage of 34% out of the total 1800 annotations (figure 1). It was followed by "Unbiased" (27%), "Biased against Israel" (16%), "Unclear" (8%), "Biased against both Palestine and Israel" (6%), "Biased against Others" (6%), and "Not Applicable" (3%).

Our guidelines set out to explore if there was a method to detect possible bias amidst a conflict of high sensitivity and relatively unfair coverage in non-Arabic news. The keywords we decided on within the guidelines included words of reference to two main actors in the conflict: the Israeli forces and the Palestinian movement. These words and the entities or events they refer to can be found in table (2). We analyze the collocation frequencies of each of the words provided and assign them bias labels, which is further explored in the next section. The guidelines helped reached an F1 centrality score of 33.1 across all other participating teams' Main batches 1 and 2. This score was the highest in terms of Bias annotation consistency.

## 5 Discussion

Thonet et al. (2016) brought up the wording surrounding the Israel-Palestine case within online articles, judging how supporters of each side use targeted words to express bias. We noted through corpus analysis that most collocations are emotionally charged and contributed to shaping the readers' views on some entities, such as the usage of "terrorists" to refer to Palestinians or "occupation" to refer to Israelis. Some words like "massacre" required context of specific events to properly identify as bias against either entity. For example, pairing it with "October" caused bias against Palestine, while with "genocide" or "Gaza" it would lead to bias against Israel. Some seemingly normal words were used almost exclusively to refer negatively to both entities or their actions, such as "captors" and "kidnap" to refer to Palestinian fighters, or "settlers"

| Label | Example |
|---|---|
| Unbiased | “Israel and Hamas have announced a four-day truce deal. Here’s a rundown of what’s happened in Gaza since October 7 ” |
| Biased against Palestine | “Documents left behind by terrorists in Kibbutz Nir Oz reveal the orderliness of the operation, the degree of the planning and the extent of the intelligence information in the possession of Hamas” |
| Biased against Israel | ““They bombed us.” This Red Crescent paramedic testifies to bombings that occurred while he was transferring civilians to the south of the Gaza Strip.” |
| Biased against both Palestine and Israel | “Massive destruction in the neighborhoods of Gaza City after fifty days of war between Hamas and Israel” |
| Biased against others | “The war between Israel and Hamas Iran’s allies in the Middle East are on alert as the war rages Follow the latest developments around the clock via the link” |
| Unclear | “NYU student Ryna Workman, who says they were denied a job following their controversial comments on the Israel-Hamas conflict, speaks with LinseyDavis.” |
| Not Applicable | “Ipsos Barometer – “Le Point”: Mélenchon in free fall” |

Table 1: Example headlines corresponding to each label in the FIGNEWS dataset.

| | |
|---|---|
| Anti-Palestine sentiment | Hamas, terrorists, massacre, attack, october, jihad, terror, isis, antisemitic, captors, kidnap, bringthemhome, israelunderattack |
| Anti-Israel sentiment | zion, settler, occupy, occupation, bombardment, siege, genocide, aggression, displace, ceasefirenow, freepalestine, gazaunderattack |
| Against other parties | houthi, hezbollah, pakistan, macron, iran, yemen, iraq, icc, icj, un, |

Table 2: Example tokens with possible negative sentiment against the main entities.

and “displace” to refer to Israeli forces. "Hamas-run" is an implicit case of anti-Palestinian bias, acting as a "dog-whistle" to a claimed non-autonomy of the Palestinian state.

Considering the social media element of this news dataset, hashtags played a major role in bias detection, with “ceasefirenow” or “freepalestine” indicating pro-Palestinian tendencies, and “bringthemhome” or “israelfightsterror” being pro-Israeli sentiment. Within the Hindi sections of the batches, a significant number of posts used an abundance of hashtags which often contradicted themselves on sentiment, causing more difficulty in bias detection, and violating Grice’s pragmatic maxim of quantity (Benamara et al., 2018).

We attempt to classify the type of media bias in the dataset according to D’Alessio and Allen (2000), where “gatekeeping bias” is the picking of specific stories and omitting others, “coverage bias” is the allocated space for each side of the story, and “statement bias” is the deliberate inclusion of opinion by journalists into reporting. We decided on “gatekeeping” and “statement bias” as the two major culprits of biased reporting during the Israeli war on Gaza. Within the corpus there were two major factors that determined potential bias on the document level, which are news outlet source and language source. This approach relates to most previous bias detection efforts as mentioned before. While we attempted to analyze token level instances to detect bias, we found it mostly correlated with its source. Source identification continues to be an important contextual supplement to the process of bias identification and creation of bias annotation guidelines.

## 6 Conclusion

We tackled the task with a linguistic perspective in hopes of offering a better understanding of the pragmatics that surround bias detection. Due to the task’s underlying complexity even to people, having machines perform statistical classifications on such sensitive matters may carry turbulence and misguidance. Language inference is closely connected to adherence to cultural and world context. While the original notion of journalism is objectivity, that becomes increasingly difficult as the news

covered becomes more sensitive to every party involved. We found that the polarity of the conflict has generated keywords which have a high possibility of better bias detection, yet a headline may be further flawed by contradictory hit word injection, or in the case of social media, by inclusion of irrelevant hashtags.

## References


Ali M Abushbak, Tawseef Majeed, and Atul Sinha. Instagram, censorship and civilian activism: The digital presence of the israel-palestine conflict narratives.

Iyad Al-Agha and Osama Abu-Dahrooj. 2019. Multilevel analysis of political sentiments using twitter data: A case study of the palestinian-israeli conflict. *Jordanian Journal of Computers and Information Technology*, 5(3).

Mohammed Alashqar. 2024. *The Role of Language in Framing the Israeli-Palestinian Conflict on Twitter: The Escalation of Violence in Gaza in May 2021 as a Case Study*. Ph.D. thesis, Ghent University.

Abeer Alhossary, Ihab Ahmed Awais, and Sujod Awais. 2023. Examining israeli media targeting arab and muslim audiences: A content analysis of the 'israel speaks arabic'facebook page. *FWU Journal of Social Sciences*, 17(4):66–79.

Farah Benamara, Diana Inkpen, and Maite Taboada. 2018. Introduction to the special issue on language in social media: exploiting discourse and other contextual information. *Computational Linguistics*, 44(4):663–681.

Alba Bonet-Jover, Robiert Sepúlveda-Torres, Estela Saquete, Patricio Martínez-Barco, and Mario Nieto-Pérez. 2023. Run-as: a novel approach to annotate news reliability for disinformation detection. *Language Resources and Evaluation*, pages 1–31.

Dimitrios Bountouridis, Mykola Makhortykh, Emily Sullivan, Jaron Harambam, Nava Tintarev, and Claudia Hauff. 2019. Annotating credibility: identifying and mitigating bias in credibility datasets. In *ROME 2019-Workshop on Reducing Online Misinformation Exposure*.

D D'Alessio and M Allen. 2000. Media bias in presidential elections: a meta-analysis. *Journal of Communication*, 50(4):133–156.

Jason Deegan, John Hogan, Sharon Feeney, and Brendan O'Rourke. 2018. The self and other: portraying israeli and palestinian identities on twitter.

Ullrich KH Ecker, Stephan Lewandowsky, Ee Pin Chang, and Rekha Pillai. 2014. The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied*, 20(4):323.

Houda Elmimouni, Yarden Skop, Norah Abokhodair, Sarah Rüller, Konstantin Aal, Anne Weibert, Adel Al-Dawood, Volker Wulf, and Peter Tolmie. 2024. Shielding or silencing?: An investigation into content moderation during the sheikh jarrah crisis. *Proceedings of the ACM on Human-Computer Interaction*, 8(GROUP):1–21.

Rita Finkbeiner. 2024. The pragmatics of headlines. central issues and future research avenues.

Tarleton Gillespie. 2022. Do not recommend? reduction as a form of content moderation. *Social Media+ Society*, 8(3):20563051221117552.

Stefan Th Gries and Andrea L Berez. 2017. Linguistic annotation in/for corpus linguistics. *Handbook of linguistic annotation*, pages 379–409.

Felix Hamborg, Karsten Donnay, and Bela Gipp. 2021. Towards target-dependent sentiment classification in news articles. In *Diversity, Divergence, Dialogue: 16th International Conference, iConference 2021, Beijing, China, March 17–31, 2021, Proceedings, Part II 16*, pages 156–166. Springer.

Christoph Hube and Besnik Fetahu. 2018. Detecting biased statements in wikipedia. In *Companion proceedings of the the web conference 2018*, pages 1779–1786.

Sora Lim, Adam Jatowt, Michael Färber, and Masatoshi Yoshikawa. 2020. Annotating and analyzing biased sentences in news articles using crowdsourcing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1478–1484.

Sora Lim, Adam Jatowt, and Masatoshi Yoshikawa. 2018. Understanding characteristics of biased sentences in news articles. In *CIKM workshops*, pages 121–128.

Tala Majzoub. 2021. Framing what's breaking: Empirical analysis of al jazeera and al arabiya twitter coverage of the gaza-israel conflict'. *Arab reform initiative*.

Alireza Mohammadinodooshan and Niklas Carlsson. 2023. Effects of political bias and reliability on temporal user engagement with news articles shared on facebook. In *International Conference on Passive and Active Network Measurement*, pages 160–187. Springer.

Sophia Ricketts. 2024. Utilizing modern media coverage of the israel palestine conflict to teach media literacy.

Maija Sevón. 2020. Frame change due to policy change: A corpus study of the changing frame of israel in the us media after jerusalem was recognized as capital.

Timo Spinde, Elisabeth Richter, Martin Wessel, Juhi Kulshrestha, and Karsten Donnay. 2023. What do twitter comments tell about news article bias? assessing the impact of news article bias on its perception on twitter. *Online Social Networks and Media*, 37:100264.

Timo Spinde, Lada Rudnitckaia, Felix Hamborg, and Bela Gipp. 2021. Identification of biased terms in news articles by comparison of outlet-specific word embeddings. In *Diversity, Divergence, Dialogue: 16th International Conference, iConference 2021, Beijing, China, March 17–31, 2021, Proceedings, Part II 16*, pages 215–224. Springer.

Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, and Karen Pinel-Sauvagnat. 2016. Vodum: a topic model unifying viewpoint, topic and opinion discovery. In *Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38*, pages 533–545. Springer.

Wajdi Zaghouani, Mustafa Jarrar, Nizar Habash, Houda Bouamor, Imed Zitouni, Mona Diab, Samhaa R. El-Beltagy, and Muhammed Raed AbuOdeh, editors. 2024. *The FIGNEWS Shared Task on News Media Narratives*. Association for Computational Linguistics, Bangkok, Thailand.

## A Appendix

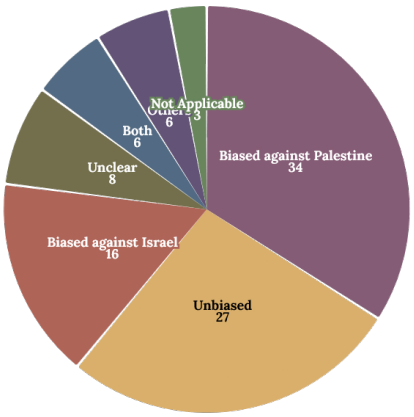


Figure 1: Percentages of labels out of a total 1800 annotations.

| Word | Collocate | Col. frequency | Log dice |
|---|---|---|---|
| attack | hamas | 121 | 11.1 |
| terrorist | hamas | 103 | 10.9 |
| occupation | israeli | 11 | 9.0 |
| bombardment | israeli | 9 | 8.9 |
| aggression | israeli | 9 | 8.8 |
| massacre | hamas | 21 | 8.7 |
| settlers | israeli | 8 | 8.7 |

Table 3: Most frequent collocations to refer to Israeli or Palestinian entities.

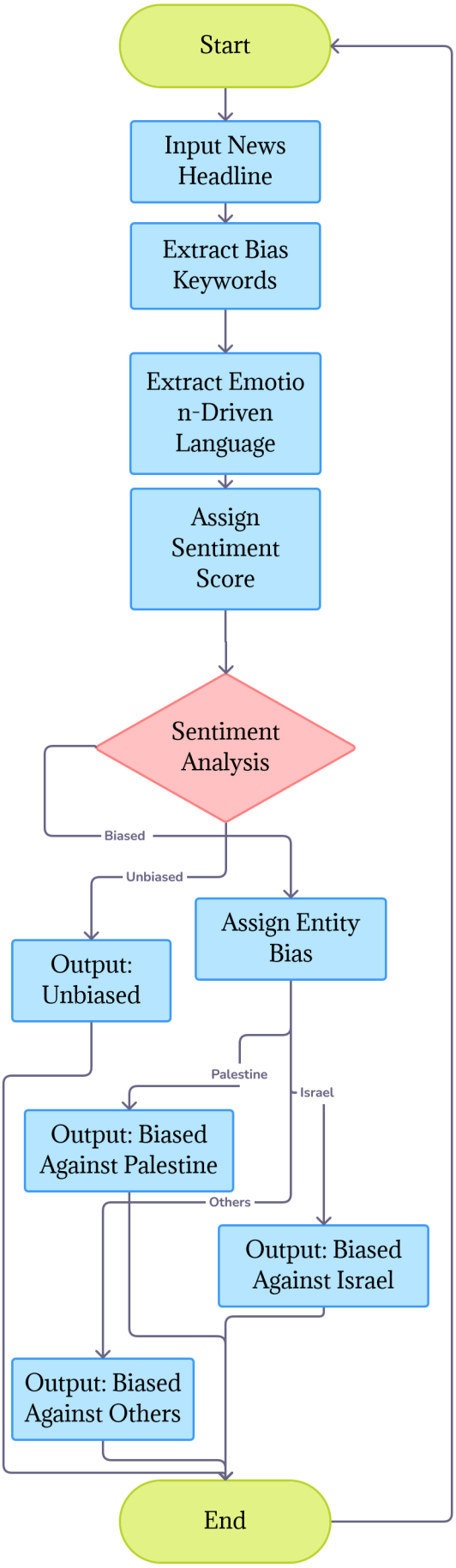


Figure 2: Theoretical system diagram for simple bias classification.