UoT1 at FIGNEWS 2024 Shared Task: Labeling News Bias

Abdusalam F A Nwesri and **Mai M. Elbaabaa**
and **Fatima A. Ben Lashihar** and **Fatma Ali Alalos**
University of Tripoli
Tripoli - Libya
{a.nwesri,m.elbaabaa,f.benlashihar,f.alalus}@uot.edu.ly

Abstract

This paper outlines the University of Tripoli's initiative in creating annotation guidelines to detect bias in news articles concerning the Palestinian-Israeli conflict. Our team participated in the Framing of Israeli Gaza News Media Narrative (FIGNEWS 2024) shared task. We developed annotation guidelines to label bias in news articles. Using those guidelines we managed to annotate 3,900 articles with the aid of our custom-developed annotation tool. Among 16 participating teams, we scored 48.7 on the macro F1 measure in the quality track in which we ranked 4th. In the centrality track we were ranked at the 6th position using the macro F1 avg measure, however, we achieved the 4th best kappa coefficient. Our bias annotation guidelines was ranked in the 9th position.

1 Introduction

Framing can be defined as the process of selecting certain elements of perceived reality and constructing a narrative that emphasizes their connections to promote a specific explanation (Entman, 2007). News framing involves emphasizing certain aspects of specific issues in news reports to encourage a particular interpretation.

While recent NLP studies have examined framing in English news, there has been limited exploration of how this analysis can be applied to other languages and in a multi-label context (Akyürek et al., 2020). This paper is a contribution to the creative Shared Task on News Media Narratives (Framing the Israel War on Gaza)(Zaghouani et al., 2024) to explore into the complex intricacies of bias and double standards evident in news coverage. The main goal is to create a common database for thorough annotation across different levels, shaping annotation rules based on the various and sometimes conflicting discussions on this delicate subject. This effort aims to highlight both difficulties and positive aspects within the data and to encourage a cooperative environment. By dissecting news articles in various languages, this joint endeavor aims to uncover the underlying biases and propaganda, encouraging a collective investigation into the media narratives surrounding this significant historical period (Zaghouani et al., 2024).

Text annotation is a crucial activity across various fields that involves several essential steps includes: formatting the text, defining annotation categories, organizing large volumes of documents for annotation, assessing inter-annotator agreement on the same documents, and preparing annotated documents in formats compatible with NLP software (Grosman et al., 2020). In Natural Language Processing (NLP) research, text annotation plays a fundamental role in generating datasets used to train and assess automated techniques (Spinde et al., 2021b). In the realm of news reporting, it's crucial for both writers and audiences to uphold principles of fairness, accuracy, and impartiality by presenting a balanced array of viewpoints. Nevertheless, bias in news articles has emerged as a significant concern. Despite numerous news outlets asserting their commitment to objectivity, individual sources often harbor distinct perspectives on societal, political, and other subjects. Additionally, the imperative to engage readers to ensure profitability can inadvertently foster biased reporting practices, potentially yielding harmful outcomes(Lim et al., 2018).

This task focuses on examining how journalists' reporting of important events can shape consumers' views on political matters (Spinde et al., 2021b). The inherent bias of news media is a fundamental flaw in the entire news production process, encompassing stages from gathering to writing and editing. At each step, the subjective judgments of producers and external pressures from owners and advertisers likely influence the news content. These influences often lead to diverse framings of reality, as producers emphasize different aspects, employ varied tones, and present stories in distinct styles.

Proceedings of The Second Arabic Natural Language Processing Conference, pages 567–572
August 16, 2024 ©2024 Association for Computational Linguistics

Consequently, this bias exacerbates political polarization and contributes to misunderstandings of crucial issues. It becomes challenging for readers to attain a balanced perspective on realities, especially when they primarily access news through limited channels, often just one (Park et al., 2009). Thus, recognizing and addressing news bias is crucial, not only for informing readers but also for guiding authors in maintaining a neutral writing style (Lim et al., 2020).

The importance of detecting bias in media news has been recognized and studied by several researchers(Vaagan et al., 2010; Al-Sarraj and Lubbad, 2018; Morstatter et al., 2018; Varacheva and Gherghina, 2018). Advanced bias detection techniques rely on artificial intelligence and machine learning (Budak et al., 2016; Spinde et al., 2021b), thus requiring a well-annotated datasets to train AI models. existing datasets often suffer from limitations such as lack of representativeness, insufficient diversity, and inadequate information about annotator characteristics. Addressing these challenges is crucial for contextualizing and interpreting bias annotations accurately. Previous research efforts have contributed to the development of datasets for bias detection, such as those by (Lim et al., 2018, 2020), (Baumer et al., 2015), (Hamborg et al., 2019), and (Fan et al., 2019). However, these datasets often have limitations such as small topic coverage, focus on specific framing effects, one-language news source, or lack of annotations at the word level. Spinde et al. (2021a) addressed the data gap by presenting the MBIC (Media Bias Including Characteristics) dataset, which contains 1,700 statements annotated by ten annotators each. This dataset represents various instances of media bias, including framing and epistemological bias, and provides labels for bias identification at both the word and sentence levels. Notably, MBIC is the first dataset to include detailed information about annotator characteristics and their individual backgrounds, enhancing the reliability and interpretability of bias annotations. In the context of Israeli-Gaza war, there exist few reliable dataset that can be used to train AI models effectively. Most existing datasets are sourced from one language and use two to three annotation labels.

The co-authors of this paper focus their effort in establishing comprehensive guidelines for annotating biased news reports concerning the current conflict in Gaza as pointed out in FIGNEWS 2024 shared task(Zaghouani et al., 2024).

section 2 describes the process of annotation including the annotators training and team coordination, Section 3 discusses the team contributions and Section 4 provides some concluding remarks and points to future work.

2 Annotation Methodology

2.1 The Task

The FIGNEWS 2024 shared task aim is to determine bias and propaganda within news articles in multiple languages. Participant teams have to develop guidelines to annotate news articles about the 2023-2024 war between Gaza and Israel. A dataset of 15000 news articles was provided by the organizers. It has been divided into 15 batches. Each batch contains 1000 news articles collected from different sources in five languages namely: English, Arabic, Hebrew, French, and Hindi. Each batch has been further divided into a 100-subset called Inter Annotation Agreement (IAA), and a 900-subsets called (MAIN). The IAA subset has to be annotated by all Annotators in the team, while the 900-subset can be annotated collaboratively by the team members. Each team is required to annotate a minimum of two batches to enter the shared task. All articles are machine translated into English and Arabic, annotators should specify the language they are using to read articles they are annotating.

> "Teams must provide well-documented annotation guidelines including examples, and must provide inter-annotator agreement (IAA) numbers for at least 200 posts (40 from each language) from Batch 1 and Batch 2. We expect the IAA to be competitive (e.g. Cohen Kappa of 0.6+) in the target space. The best guidelines will be selected by the organizers."

Our Team choose to participate in the bias task. This document will present guidelines for annotating bias in the provided subset.

2.2 Team Composition and Training

Eight academic staff members from the the department of software engineering, faculty of information technology at the University of Tripoli participated in this task. Participants are native-Arabic speakers with strong command of English language. Two of the participants are experts in the annotations process while the rest had no experience about the task. Participants were divided into two teams

lead by the two senior experts. Initial meeting were made to introduce participants about the task, the dataset and the annotation process. At the beginning, the two teams were interested in participating in Bias and Propaganda annotations, however, later on a decision made that each team participates in one annotation task. In our case, we participated in the Bias annotation which is the focus of this paper.

The team was initially trained on how to use the annotation sheet provided by the organizers, however, using the sheet to read long articles was cumbersome. As a result, an in-house annotation tool specifically developed to facilitate the annotation process where articles are presented based on the language chosen by teams for annotation, with Arabic being the focus in their case. Annotators can review the Arabic version of the article and annotate it using the specified task, selecting the most appropriate label for each article. The focus of training was on familiarize inexperienced team members on the task and to reach a consensus on annotation labels.

2.3 Annotation Guidelines

Initially adopted labels provided by the organizers however, we added two more labels after the first round of training. The labels used to annotate bias in the news articles are:

- Unbiased
- Biased against Palestine
- Biased against Israel
- Biased against both Palestine and Israel
- Biased against others
- Unclear
- Not Applicable

2.4 Annotation Process

The annotation process started by asking annotators to use the initial labels and examples found in the shared task detailed page [1]. Each annotator has to annotate the first batch of IAA articles. They were asked to write down comments and to discuss any differences in annotations in a second meeting. They used the in-house tool to view their annotation agreement visually. They were also asked to raise any questions regarding ambiguous articles. Once the first round was done, the team discussed difficulties they faced in annotating articles, especially using the "Unclear" label. Another issue was raised is that there are articles which cannot be labeled as "Unclear" as it serves one side in the conflict than the other. For example: the article “Golani Ba: The fighters in the Hamas parliament building in Gaza. May God protect them!” cannot be labeled as "Bias against Palestine" nor "Unclear", however it is biased in favor of Israel. Similarly, the article "Erdogan: Hamas is not a terrorist organization. Rather, they struggle to defend their people and their land. They are a group of mujahideen and a liberation organization." is not "Biased Against Israel" nor "Unclear" but it is in favour of Palestine. As a result of the first annotation round, two new labels were added:

- "Biased in Favour of Palestine"
- "Biased in Favour of Israel"

Another session is made to annotate a group of articles together. This clarifies several differences found in the first round. For example, it has been agreed on the exact meaning of "Not-applicable" and "Unclear" labels.

2.5 Inter-Annotator Agreement (IAA) Analysis

During the annotation process, annotators could visually view the annotators agreement through the dashboard of the annotation tool. We did not use any measures to compute team agreement, but both annotators and team leader were able to view the agreement in any batch using the dashboard of the annotation tool as shown in Figure 1.

Annotators are left with their choice of annotations, but they can go back and update their annotation label. Based on the initial annotation results, there is a moderate disagreement between annotations in the IAA batches. Despite using clear examples, annotators in many cases do not agree on labeling articles. The most agreed articles between annotators are those labeled "Unbiased" followed by "Biased against Palestine" labels. The number of annotations made by team members are shown in the dashboard and can be viewed by all. This made as a motivation to compete in annotating articles.

Annotating the "MAIN" subset is made by the whole team. The annotation tool has been adjusted

[1] https://sites.google.com/view/fignews/shared-task-details

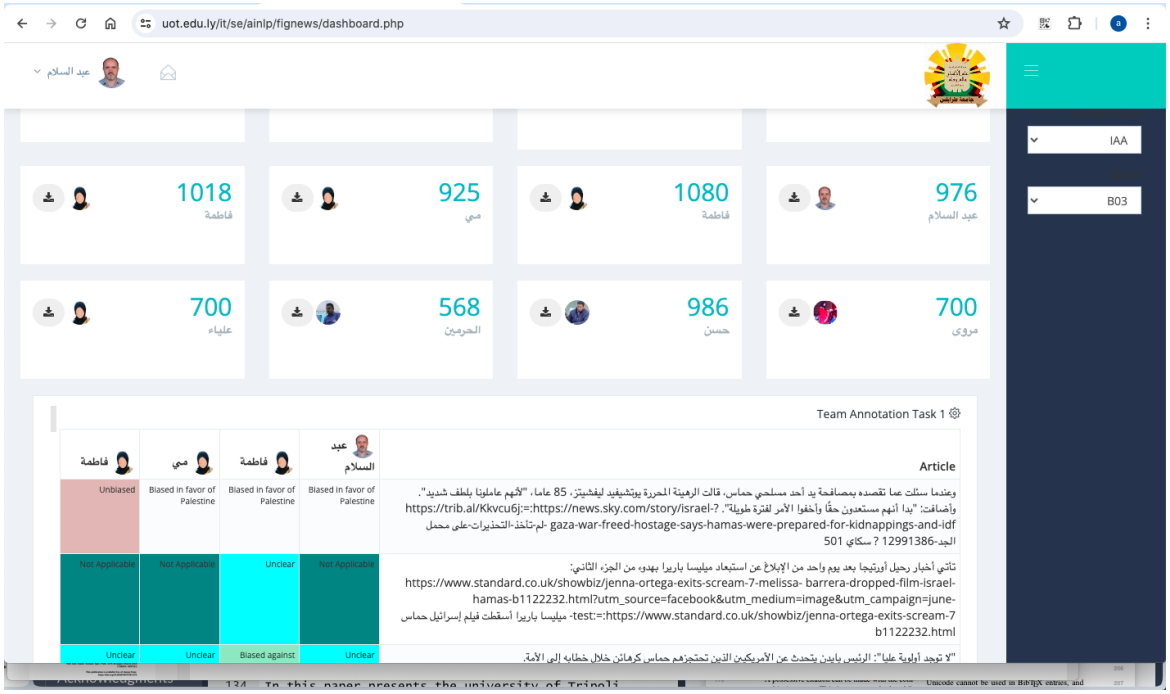

Figure 1: IAA agreement visual view: Annotators can view team annotations on a specific batch.

to show only un-annotated articles to team members and automatically hide articles once they are annotated.

The team managed to completed annotating the first three batches. final annotations were exported by the tool and copied to the google sheet prepared specifically by the organizers for the task.

3 Results

The shared task included 23 participating teams. only 17 teams qualified by finishing the required batches. Results were published by the organizers and made public[2]. Different measures were calculated to assess the quantity, quality, and centrality (consistency). Our rank in the three measures ranges from 4th to 8th position across all measures. In the quantity track, we were placed at the 7th position with 3900 points. We believe that teams would have been ranked better in this measure if effort per an annotator is considered as team size ranges between 2 to 21 annotator. In the quality track, we were ranked 4th using the macro F1 average measure and in the centrality track we were ranked 4th using the kappa measure. Our average rank is 6 cross all ranks. If applied to all teams, we could have been ranked in the second position.

[2] https://docs.google.com/spreadsheets/d/1HZ_fGaJkbZvEdFM8Gwghd2mn-6mE8u2cXYCcBwfp9ao/edit?gid=1947761727#gid=1947761727

4 Discussion

Annotating bias in news articles is a complex and meticulous task that requires careful execution. Despite the design of labels and provision of examples, annotators encountered several challenges that led to mistakes in labeling articles correctly.

Firstly, the annotation process itself is tedious, demanding both patience and concentration. The progress of annotators was notably faster during the initial rounds compared to the subsequent rounds, indicating the need for sustained effort and focus over time.

Secondly, the reliance on translation introduced additional difficulties. For example, the term "Hamas," which refers to a leading Palestinian organization in the conflict, also means "enthusiasm" in Arabic. This dual meaning led to misleading annotations in some cases. For instance, in an article translated from English titled "Cheering Hamas because a released hostage smiled is a pretty low bar for humanity | Haaretz Today,"

إن الهتاف بحماس بسبب ابتسامة الرهينة المفرج عنها هو مستوى منخفض جدًا للإنسانية ...

the term "Hamas" was misinterpreted as "enthusiasm" in the Arabic version, resulting in annotation discrepancies.

To address some of these challenges, we initially added two more labels to better capture the nuances

in the text. However, we were later instructed by the organizers to remove these additional labels and re-annotate the articles using the original labels specified in the shared task. This re-annotation process underscored the importance of adhering to standardized guidelines to maintain consistency and accuracy in the annotations

5 Conclusions

The University of Tripoli's initiative to develop guidelines for annotating bias in news articles about the Palestinian-Israeli conflict has yielded significant progress in understanding and detecting media bias. By participating in the FIGNEWS 2024 shared task, we successfully annotated over 3,000 articles and created an annotation tool that non-experts can use to identify bias effectively. Our findings underscore the complexities involved in annotating bias, especially given the challenges posed by multilingual datasets and translation nuances. This effort highlights the necessity for clear, detailed guidelines to ensure accuracy and consistency in annotations. The collaborative aspect of FIGNEWS 2024 has been instrumental in achieving reliable inter-annotator agreement, which is crucial for the validity of the annotated data. This project makes a valuable contribution to Natural Language Processing by offering new resources and insights into media bias detection. It lays the groundwork for future research and innovation, aimed at promoting balanced and fair news reporting. Moving forward, we plan to further refine and expand our guidelines and tools, enhancing the capacity of researchers and practitioners to identify and mitigate bias in media narratives.

References

Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8614–8624, Online. Association for Computational Linguistics.

Wael Al-Sarraj and Mohammed Lubbad. 2018. Bias detection of palestinian/israeli conflict in western media: A sentiment analysis experimental study. pages 98–103.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, editor = "Mihalcea Rada Gay, Geri", Joyce Chai, and Anoop Sarkar. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482, Denver, Colorado. Association for Computational Linguistics.

Ceren Budak, Sharad Goel, and Justin Rao. 2016. Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80:250–271.

Robert M Entman. 2007. Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173.

Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. In plain sight: Media bias through the lens of factual reporting. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6343–6349, Hong Kong, China. Association for Computational Linguistics.

Jonatas Grosman, Pedro Thompson, Ariane Rodrigues, Guilherme Schardong, Simone Barbosa, and Hélio Lopes. 2020. Eras: Improving the quality control in the annotation process for natural language processing tasks. *Information Systems*, 93:101553.

Felix Hamborg, Anastasia Zhukova, and Bela Gipp. 2019. Automated identification of media bias by word choice and labeling in news articles.

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, Marseille, France. European Language Resources Association.

Sora Lim, Adam Jatowt, and Masatoshi Yoshikawa. 2018. Understanding characteristics of biased sentences in news articles. In *CIKM Workshops*, volume 2482 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen Corman, and Huan Liu. 2018. Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1:1–18.

Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. Newscube: Delivering multiple aspects of news to mitigate media bias. pages 443–452.

Timo Spinde, Jan-David Krieger, Manuel Plank, and Bela Gipp. 2021a. Towards A reliable ground-truth for biased language detection. *CoRR*, abs/2112.07421.

Timo Spinde, Kanishka Sinha, Norman Meuschke, and Bela Gipp. 2021b. TASSY - A text annotation survey system. *CoRR*, abs/2112.07391.

Robert Vaagan, Frøydis Johannessen, and Marie Walsøe. 2010. Tv news and "white voices": Dagsrevyen's coverage of the gaza war. *javnost - the Public*, 17(3):39–56.

Tatiana Varacheva and Sergiu Gherghina. 2018. Neutral or biased? the presentation of the kyrgyzstan and egypt uprisings by ria novosti. *Europe-Asia Studies*, 70:1–23.

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