The FIGNEWS Shared Task on News Media Narratives

Wajdi Zaghouani,[1] **Mustafa Jarrar,**[2] **Nizar Habash,**[3] **Houda Bouamor,**[4]
Imed Zitouni,[5] **Mona Diab,**[6] **Samhaa R. El-Beltagy,**[7] **Muhammed AbuOdeh**[3]
[1]Northwestern University in Qatar, Education City, Doha, Qatar [2]Birzeit University, Palestine
[3]New York University Abu Dhabi, UAE [4]Carnegie Mellon University, Qatar
[5]Google, USA [6]Carnegie Mellon University, USA [7]Newgiza University, Egypt
wajdi.zaghouani@northwestern.edu, mjarrar@birzeit.edu, nizar.habash@nyu.edu
mdiab@cs.cmu.edu, samhaa@computer.org, mra9047@nyu.edu

Abstract

We present an overview of the FIGNEWS shared task, organized as part of the ArabicNLP 2024 conference co-located with ACL 2024. The shared task addresses bias and propaganda annotation in multilingual news posts. We focus on the early days of the Israel War on Gaza as a case study.[1] The task aims to foster collaboration in developing annotation guidelines for subjective tasks by creating frameworks for analyzing diverse narratives highlighting potential bias and propaganda. In a spirit of fostering and encouraging diversity, we address the problem from a multilingual perspective, namely within five languages: English, French, Arabic, Hebrew, and Hindi. A total of 17 teams participated in two annotation subtasks: bias (16 teams) and propaganda (6 teams). The teams competed in four evaluation tracks: guidelines development, annotation quality, annotation quantity, and consistency. Collectively, the teams produced $129,800$ data points. Key findings and implications for the field are discussed.

1 Introduction

The FIGNEWS 2024 shared task,[2,3] henceforth FIGNEWS, addresses the critical need for analyzing bias and propaganda in multilingual news discourse surrounding the Israel War on Gaza. This task aligns with the NLP community's growing efforts to create datasets and guidelines for complex opinion analysis through collaborative shared tasks and datathons. Such a meta-task allows for the exploration of various annotation frameworks in particular for complex and challenging subjective tasks. FIGNEWS focuses on a diverse multilingual corpus, emphasizing the development of guidelines with rich examples, while fostering a research-oriented collaborative environment. By simultaneously examining multiple languages, comparing and contrasting various narratives, FIGNEWS aims to unravel the layers of possible bias and propaganda within news articles with alternative media narratives. This is especially critical at times of any war or conflict. The media's portrayal of events during such events has significant implications on public perception, policy-making, and international relations. By addressing bias and propaganda, we hope to illuminate the varied ways in which news can shape, and sometimes distort, public understanding of complex geopolitical events. This initiative seeks to explore diverse perspectives, cultures, and languages, thereby fostering a comprehensive understanding of events through the lens of major news outlets across the globe.

Developing guidelines for complex data is a challenging task. The problem is exacerbated when the data is contemporaneous hence the annotators and the task organizers might have a stance on the subject matter. FIGNEWS is our attempt at addressing the creation of annotation guidelines, addressing what relevant best practices should be. We use the Israel War on Gaza as a use case to highlight some of these aspects. To that end, we curate a shared corpus for comprehensive annotation across various layers, crafting annotation guidelines shaped by the diverse range of conflicting discourses around this sensitive topic. This endeavor facilitates the development of robust methodologies and metrics for detecting and analyzing bias and propaganda, which are crucial for ensuring fair and accurate media reporting. The curated corpus, along with meticulously developed guidelines and annotations, will serve as a valuable resource for future research in NLP and related fields.

This initiative also seeks to bring to light both challenges and commendable aspects within the data, fostering a collaborative community that can

[1]FIGNEWS: *Framing the Israel War on Gaza News*.
[2]Website: https://sites.google.com/view/fignews
[3]Code and Data: https://github.com/CAMeL-Lab/FIGNEWS-2024 & https://huggingface.co/datasets/CAMeL-Lab/FIGNEWS-2024

Proceedings of The Second Arabic Natural Language Processing Conference, pages 530–547
August 16, 2024

learn from each other's approaches and findings. We believe that a collaborative, research-oriented environment is essential for tackling the intricate task of bias and propaganda detection.

The FIGNEWS shared task thus represents a significant step forward in the field of data annotation, media analysis and NLP. A step towards an annotation science. It not only provides a platform for examining critical issues of bias and propaganda but also promotes the development of best practices in data annotation and analysis. Through this shared task, we aim to contribute to the broader goal of improving media literacy and fostering a more informed and critically engaged public.

2 Related Work

Propaganda and bias can have far-reaching implications depending on the context and medium in which they are propagated. Polarization, conflict and injustice are but some of the side effects they can create in any context. In the context of politics, they can alter the outcomes of elections and change the face of nations (Gorenc, 2020; Maweu, 2019; Solopova et al., 2024). News media framing and narratives around wars and socio-political events have been extensively studied, with a focus on identifying bias, propaganda, offensive language detection, and diverse perspectives (Entman, 2007; Baumer et al., 2015; Fan et al., 2019; Morstatter et al., 2018; Park et al., 2009; Martino et al., 2020b; Aksenov et al., 2021; Yenkikar et al., 2022; Kameswari et al., 2020; Hong et al., 2023; Kim et al., 2023; Sharma et al., 2023; Maab et al., 2023; Rodrigo-Ginés et al., 2024; Hamad et al., 2023; Darwish et al., 2021). Several works have proposed computational approaches to detect media bias through analysis of word choice, labeling, and factual reporting (Hamborg et al., 2019; Budak et al., 2016; Vaagan et al., 2010; Varacheva and Gherghina, 2018).

Entman (2007) defines *framing* as the process of selecting certain aspects of perceived reality and constructing a narrative that emphasizes their connections to promote a specific interpretation. This foundational work highlights how news framing can influence public perception by emphasizing particular elements over others. Entman (1993) further elaborates on this concept, providing a comprehensive framework for understanding how media frames can shape political and social realities.

Several studies have developed methods and datasets to detect news bias. Budak et al. (2016) used crowdsourced content analysis to quantify media bias, while Baumer et al. (2015) compared computational approaches for detecting framing in political news. Tools like Biasly (2017) help users gauge news bias, quantifying liberal or conservative leanings.

In the context of multilingual and multi-label news framing analysis, Akyürek et al. (2020) explored the complexities of news framing across different languages. Their work is crucial for understanding how bias and framing manifest in multilingual contexts, aligning closely with the objectives of the FIGNEWS shared task. Similarly, the study by (Heppell et al., 2023) offers a valuable dataset and linguistic insights from two multilingual disinformation websites, providing a foundation for further exploration of language-specific disinformation techniques and their impact on news framing.

Recent advances in bias detection have also emphasized the importance of detailed and well-annotated datasets. Spinde et al. (2021a) introduced the MBIC dataset, which includes detailed information about annotator characteristics and provides labels for bias identification at both the word and sentence levels. This dataset represents a significant step forward in creating reliable ground-truth data for bias detection. Additionally, Spinde et al. (2021b) developed TASSY, a text annotation survey system that enhances the quality control of annotation processes in NLP tasks.

Annotation of biased language and framing in news articles has been explored using techniques like expert annotation (Al-Sarraj and Lubbad, 2018) and crowdsourcing (Lim et al., 2020, 2018). Quality control and guidelines for annotation processes in NLP tasks have also been investigated (Grosman et al., 2020; Spinde et al., 2021b). Grosman et al. (2020) developed ERAS, a system designed to enhance quality control in NLP tasks, which is particularly relevant for ensuring the reliability of annotations in bias and propaganda detection.

Further contributions to the detection of bias and propaganda include the work of Rashkin et al. (2017), who analyzed language in fake news and political fact-checking, and Barrón-Cedeno et al. (2019), who organized news based on their propaganda-prone content. These studies provide foundational techniques for the identification and analysis of biased and propaganda language in news media.

The shared task on propaganda detection at SemEval-2020, organized by Martino et al. (2020a), is directly relevant to the FIGNEWS task. This task focused on detecting propaganda techniques in news articles, providing valuable benchmarks and methodologies that can be applied to the detection of bias and propaganda in social media posts.

In the context of the Israel-related wars and conflicts, Al-Sarraj and Lubbad (2018) conducted a sentiment analysis study to detect bias in Western media coverage. This work highlights the specific challenges of detecting bias in a highly polarized and sensitive geopolitical context. Additionally, Hamad et al. (2023) introduced a new dataset designed for detecting and classifying various forms of offensive language in Hebrew on social media. Furthermore, the WojoodNER Shared Task 2024 offered a new NER dataset related to the Israeli War on Gaza called *Wojood*^{Gaza} (Jarrar et al., 2024).

Research on detecting political bias in social media includes Zhou et al. (2020), who identified bias in user-generated content, and Li et al. (2019), who surveyed sentiment analysis techniques. The Propitter corpus (Casavantes et al., 2024) was developed using distant supervision with refined annotations. Hamborg et al. (2019) advanced media bias detection by focusing on automated identification through word choice and labeling. Fan et al. (2019) analyzed factual reporting to reduce bias, stressing the need for objectivity.

Among the studies forming a strong basis for understanding bias and propaganda detection in media are the following. Baly et al. (2020) examined news media profiling using text and social media analysis, showing how news context affects bias perception. Allcott and Gentzkow (2017), Vosoughi et al. (2018), and Grinberg et al. (2019) studied fake news spread during the 2016 US election, highlighting the need for reliable information. Lazer et al. (2018) detailed fake news detection methods, while Horne and Adali (2017) emphasized high-quality data for combating misinformation. Sharma et al. (2019) surveyed fake news identification techniques. Zhou and Zafarani (2020) reviewed fake news detection theories and methods, pointing to future research opportunities.

While prior work has made notable contributions, our shared task focuses on comprehensive annotation of media narratives and framing around a specific war to uncover new insights. We build upon existing annotation frameworks (Zaghouani et al., 2016; Zaghouani and Charfi, 2018) to develop improved guidelines and foster a collaborative exploration of media discourses through a multilingual, multicultural lens.

3 Data Collection and Selection

We used the CrowdTangle[4] platform to collect Facebook posts related to the Israel War on Gaza in five languages: English, French, Arabic, Hebrew, and Hindi. Specifically, we retrieved public posts containing the keyword "Gaza"[5] from verified blue-check accounts, publicly posted between October 7, 2023, and January 31, 2024.

For each language, we collected approximately $300,000$ posts. To narrow this down to $3,000$ posts per language and have a balanced distribution, we focused on ten key moments in the first few months of the war: starting with the events of October 7, 2023 and the declaration of war, and including the bombings in Jabalia refugee camp, Al-Shifa hospital, and St. Porphyrius Church, as well as the ceasefire and hostage release, Gaza mass arrests, and the coverage of the International Court of Justice case till the end of January 2024.

For each of the ten moments, we selected the top 200 posts per language, ranked by total interaction count (likes, shares, comments, etc.). As a result, we selected $15,000$ posts with the highest interactions, covering different key moments of the War, in five languages.

We divided the $15,000$ posts into 15 batches ({`B01`, `B02`, ..., `B015`}) with $1,000$ posts in each batch. Each batch contains 200 posts per language. Additionally, each batch included 20 posts for each of the 10 key moments during the War.

Since the annotators may not speak all five languages, we provided machine translations (Google Translate) into English and Arabic to facilitate annotation across the multilingual corpus. While machine translation inevitably introduces some noise, we considered this to be reflective of the real-world reliance on such technologies.

To compare the Inter-Annotator Agreement (IAA) among all participating teams in a fair manner, we randomly selected 20 posts from each language (i.e., 100 posts, which is 10%) from each batch, totaling $1,500$ posts. All participants received the dataset in a Google Sheet file with two sheets: "Main" and "IAA". The "Main" sheet con-

[4] https://crowdtangle.com

[5] غزة, עזה, गाजा, and Gaza.

tains 13, 500 posts, while the "IAA" sheet contains 1, 500 posts. See Appendix B for a screenshot of the interface.

It is important to note that participants were provided only with the original posts and their translations. Information such as account owner, date, and total interactions were not given to any participating team (See the ethical consideration section).

4 Subtasks and Evaluation Tracks

This section presents the shared task subtasks, evaluation tracks and minimal requirement for teams to qualify. Further details are in Appendix A.

4.1 Minimal Requirements to Qualify

To qualify, each participating team **must provide full annotation guidelines** for each subtask they choose to work on; and they **must annotate at minimum Batch 1 and Batch 2**, i.e. (1,800 posts) and their designated Inter-annotator agreement subset (200 posts) for a total of 2,000 posts.

4.2 Annotation Subtasks

The shared task consists of two subtasks: Bias Annotation and Propaganda Annotation.

4.2.1 Bias Annotation Subtask

The Bias Annotation subtask involves assigning one of the following seven labels to each post: (1) **Unbiased**, (2) **Biased against Palestine**, (3) **Biased against Israel**, (4) **Biased against both Palestine and Israel**, (5) **Biased against others**, (6) **Unclear**, and (7) **Not Applicable**. Examples illustrating each label are provided in the shared task description in Appendix A.

4.2.2 Propaganda Annotation Subtask

The Propaganda Annotation subtask requires participants to classify each post into one of the following four categories: (1) **Propaganda**, (2) **Not Propaganda**, (3) **Unclear**, and (4) **Not Applicable**. Examples showcasing each category are included in the shared task description in Appendix A.

4.3 Evaluation Tracks

For each subtask, there are four evaluation tracks.

4.3.1 Guidelines Track

Participants have the freedom to design their own annotation guidelines and apply them to the shared data. The organizers will evaluate the guidelines based on an 8-point checklist, which includes items such as defining objectives, describing the task, establishing categories, providing detailed guidelines with examples, outlining the annotation process, setting quality standards, handling ambiguities, ensuring consistency, and considering ethical aspects.

The **Guidelines Score** used to determine the winners of this track is the average normalized Document Score and normalized IAA Kappa score. The Document Score is equal to the number of satisfied document checklist items, a range from 0 to 8. The IAA Kappa score of a team is the average of all pairwise IAA Kappas over team annotators per batch (same as IAA Quality Score discussed next). For both sub-scores, normalization is accomplished by dividing by the maximum value attained by any qualified team. Further details are in Appendix A.

4.3.2 IAA Quality Track

In the IAA Quality Track, teams compete based on their internal (within team) IAA Kappa scores (Cohen, 1960). In addition to the Kappa score (our primary metric), we report on a number of other useful metrics.

- **IAA Kappa** (Primary Metric) A team's IAA Kappa score is calculated as the average of all pairwise Kappa scores between team annotators for each relevant IAA batch abd subtask.
- **Accuracy (Acc)** The percentage of agreed upon data points (Bias and Propaganda)
- **Macro F1 Average** The average F1 score over all the Bias or Propaganda subtask labels over all pairs of annotators and relevant IAA batches, i.e. batches which the pairs of annotators annotated.
- **F1 Bias*** The value of the average F1 score of all Bias labels collapsed as Bias vs other.
- **F1 Prop*** The value of the average F1 score of *Propaganda* label alone.

4.3.3 Quantity Track

In the Quantity Track, teams compete based on the number of annotated data points. They must complete the batches in order and finish one batch before moving to the next.

4.3.4 Consistency Track

In the Consistency Track, teams compete based on the centrality of their annotation choices compared to all other teams. Centrality reflects the consistency of a team's annotations with those of other teams. We define a team's centrality as **Macro F1**

Team Name	System Description	Bias	Propaganda
Bias Bluff Busters	Pareti et al. (2024)	X	X
BiasGanda	Al Wardi et al. (2024)	X	
BSC-LANGTECH	Ruiz-Fernández et al. (2024)	X	
Ceasefire	Sadiah et al. (2024)	X	
DRAGON	Jafari et al. (2024)	X	
Eagles	Ean et al. (2024)	X	
Groningen Annotates Gaza	Khatib et al. (2024)	X	
JusticeLeague	Saleh et al. (2024)	X	
Narrative Navigators	Al Emadi et al. (2024)	X	X
NLPColab	Rauf et al. (2024)	X	X
Sahara Pioneers	Solla et al. (2024)		X
Sina	Duaibes et al. (2024)	X	X
SQUad	Al-Mamari and Al-Farsi (2024)	X	
The CyberEquity Lab	Helal et al. (2024)	X	X
The Guideline Specialists	Bourahouat and Amer (2024)	X	
The Lexicon Ladies	El-Ghawi et al. (2024)	X	
UoT1	Nwesri et al. (2024)	X	
	Totals	16	6

Table 1: The participating teams: names, papers, and subtasks.

Average of its Bias or Propaganda annotations (as relevant) against other teams' annotations. A more central team, with higher consistency, is one that other teams agree with more on average. We report on all the metrics mentioned in Section 4.3.2 except that we consider the annotations in Main Batch 1 and Batch 2 for this track, and only compare annotators in different teams (across teams).

5 Results

5.1 Teams

Out of the 23 teams that registered, only 17 technically qualified per the rules of the shared task, i.e., minimally provided batches 1 and 2 in Main and IAA fully. Table 1 lists the qualifying teams and the subtasks they participated in. All of the qualifying teams submitted system description papers which are included in the proceedings.

5.2 Annotators

In total, 85 annotators from 16 teams participated in the Bias subtask, and 51 annotators from 6 teams participated in the Propaganda subtask. Table 10 in Appendix C highlights the diversity among the annotators based on the information they provided voluntarily. Only half of the annotators are native Arabic speakers and close to one-third are Urdu speakers. Half are between the ages 18-24 and close to one-third 25-34. Over three-quarters identify as female. They claim many regions of origin (South Asia 33%, Levant 27%, North Africa 13%, Western Europe 11%, among others). Almost all are highly educate with close to half with Master's degree. Two thirds of the annotators come from Engineering and Technology areas of expertise.

5.3 Bias Subtask

In total, there were 85 annotators across 16 teams and they annotated together 72 Main sets and 237 IAA sets, for a total of 88,500 data points. Table 2 and Table 3 present the results on the Bias Subtask. The winners are presented in Table 5.

Guidelines Track The winners of the Bias Guidelines Track are NLPColab (1st), Eagles (2nd) and Narrative Navigators (3rd). Details on their scores and ranking are in Table 2.

IAA Quality Track The winners of the Bias IAA Quality Track are NLPColab (1st), JusticeLeague (2nd) and Sina (3rd). Details on their scores are in Table 3.

Quantity Track The winners of the Bias Quantity Track are DRAGON (1st), NLPColab (2nd) and Sina (3rd). The sum of their annotations equal to 56% of all Bias subtask annotations. Details on their scores are in Table 3.

Consistency Track The winners of the Bias Consistency Track are The Lexicon Ladies (1st), NLPColab (2nd) and Narrative Navigators (3rd). Details on their scores are in Table 3.

Observations As anticipated, within-team IAA scores significantly surpass across-team IAA, with an average absolute increase of 22.6% (**Kappa**) and 19.4% (**Macro F1 Average**). The **F1 Bias*** scores, indicating binary bias determination, show

	Team	Define Objectives	Establish Categories	Provide Examples	Outline Process	Quality Standards	Handle Ambiguities	Ensure Consistency	Training & Support	Guidelines Document Score	IAA Kappa	Guidelines Score	Final Rank
Bias	**Bias Bluff Busters**	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	7	43.3	0.7120	4
	BiasGanda	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8	31.0	0.6964	6
	BSC-LANGTECH	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	6	51.0	0.6982	5
	Ceasefire	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8	26.6	0.6685	8
	DRAGON	No	No	No	Yes	Yes	No	Yes	No	3	35.7	0.4140	16
	Eagles	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**8**	**55.5**	**0.8519**	**2**
	Groningen Annotates Gaza	Yes	Yes	Yes	No	No	No	No	No	3	43.5	0.4634	15
	JusticeLeague	Yes	Yes	Yes	Unclear	Yes	Unclear	No	No	4	64.4	0.6585	10
	Narrative Navigators	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**8**	**39.4**	**0.7497**	**3**
	NLPColab	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Unclear**	**Yes**	**Yes**	**7**	**78.8**	**0.9375**	**1**
	Sina	Yes	Yes	No	No	No	No	No	Yes	3	61.4	0.5771	14
	SQUad	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8	23.4	0.6483	11
	The CyberEquity Lab	Yes	Yes	No	Yes	No	No	Yes	Yes	5	48.1	0.6175	13
	The Guideline Specialists	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	7	28.6	0.6192	12
	The Lexicon Ladies	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7	37.2	0.6732	7
	UoT1	Yes	Yes	Yes	Yes	Yes	No	Yes	Unclear	6	44.9	0.6598	9
Propaganda	**Bias Bluff Busters**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**Yes**	**7**	**31.5**	**0.6632**	**2**
	Narrative Navigators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8	12.8	0.5913	5
	NLPColab	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Unclear**	**Yes**	**Yes**	**7**	**69.9**	**0.9375**	**1**
	Sahara Pioneers	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	7	27.9	0.6373	4
	Sina	**Yes**	**Yes**	**No**	**No**	**No**	**No**	**No**	**Yes**	**3**	**65.3**	**0.6551**	**3**
	The CyberEquity Lab	Yes	Yes	No	Yes	No	No	Yes	Yes	5	33.5	0.5521	6

Table 2: Results of the guidelines evaluation track

a wide range of disagreements within teams (average 71.3%, stdev 10.9%) and across teams (average 60.2%, stdev 6.5%). This variability reflects the inherent subjectivity and complexity of bias labeling in general.

The bias labeling task is challenging, with high IAA difficult to achieve both within and across teams. However, the high scores of top-performing teams highlight the need for meticulous attention to detail and comprehensive training.

5.4 Propaganda Subtask

In total, there were 51 annotators across 6 teams and they annotated together 36 Main sets and 89 IAA sets, for a total of 41,300 data points. Table 2 and Table 4 present the results on the Propaganda Subtask. The winners are presented in Table 5.

Guidelines Track The winners of the Propaganda Guidelines Track are NLPColab (1st), Bias Bluff Busters (2nd) and Sina (3rd). Details on their scores and ranking are in Table 2.

IAA Quality Track The winners of the Propaganda IAA Quality Track are NLPColab (1st), Sina (2nd) and The CyberEquity Lab (3rd). Details on their scores are in Table 4.

Quantity Track The winners of the Propaganda Quantity Track are NLPColab (1st), Sina (2nd) and The CyberEquity Lab (3rd). The sum of their annotations equal to 82.1% of all Propaganda subtask annotations. Details on their scores are in Table 4.

Consistency Track The winners of the Propaganda Consistency Track are NLPColab (1st), Bias Bluff Busters (2nd) and Sahara Pioneers and The CyberEquity Lab (tied 3rd). Details on their scores are in Table 4.

Observations As anticipated, within-team IAA scores significantly surpass across-team IAA, with an average absolute increase of over 21.1% (**Kappa**) and 18.9% (**Macro F1 Average**). The **F1 Prop*** scores show a wide range of disagreements within teams (average 66.5%, stdev 15.4%) and across teams (average 55.5%, stdev 6.1%). This variability reflects the inherent subjectivity and complexity of this task.

Like Bias labeling, the Propaganda labeling task is quite demanding. Although it has a smaller number of labels, we see comparable patterns in terms of performance across a number of metrics.

	Quantity				Quality				Centrality			
		Batches		Total	IAA Within Team				Main B1+B2 Across Teams			
Team	Anno #	Main	IAA	Data Points	Kappa	Acc	Macro F1 Avg	F1 Bias*	Kappa	Acc	Macro F1 Avg	F1 Bias*
Bias Bluff Busters	4	2	8	2,600	43.3	56.3	48.5	69.3	14.4	28.7	21.0	61.7
BiasGanda	4	2	4	2,200	31.0	51.5	31.5	66.4	26.0	43.6	29.5	64.2
BSC-LANGTECH	2	2	4	2,200	51.0	65.5	39.8	81.5	26.5	46.6	29.2	60.2
Ceasefire	3	2	6	2,400	26.6	46.3	29.3	61.2	24.2	42.0	27.2	66.0
DRAGON	4	15	60	**19,500**	35.7	75.5	41.0	43.2	19.7	41.1	21.9	59.7
Eagles	4	2	8	2,600	55.5	75.4	48.4	68.5	25.6	46.0	25.4	55.3
Groningen Annotates Gaza	7	2	14	3,200	43.5	56.8	39.8	69.9	25.3	28.9	25.7	56.4
JusticeLeague	3	2	6	2,400	**64.4**	83.7	63.8	73.6	19.9	43.3	19.6	46.5
Narrative Navigators	7	2	4	2,200	39.4	56.5	45.5	70.5	28.0	44.5	**30.5**	66.6
NLPColab	21	15	30	**16,500**	**78.8**	85.3	76.1	94.3	27.7	42.4	**30.8**	67.3
Sina	10	12	24	**13,200**	**61.4**	81.4	55.4	75.1	11.8	38.7	17.2	48.1
SQUad	2	4	8	4,400	23.4	40.8	27.2	66.8	-0.2	8.5	5.8	56.2
The CyberEquity Lab	5	3	15	4,200	48.1	71.6	39.5	70.5	25.0	46.5	24.1	58.0
The Guideline Specialists	2	2	30	4,800	28.6	51.3	34.9	84.4	21.0	36.8	24.5	66.2
The Lexicon Ladies	3	2	4	2,200	37.2	53.0	35.4	73.4	29.1	44.1	**33.1**	66.3
UoT1	4	3	12	3,900	44.9	58.2	48.7	71.5	26.8	42.5	29.0	64.6
Average	5.3	4.5	14.8	5,531	44.5	63.1	44.1	71.3	21.9	39.0	24.7	60.2
Total	85	72	237	88,500								

Table 3: Results of the Bias subtask.

	Quantity				Quality				Centrality			
		Batches		Total	IAA Within Team				Main B1+B2 Across Teams			
Team	Annot #	Main	IAA	Data Points	Kappa	Acc	Macro F1 Avg	F1 Prop*	Kappa	Acc	Macro F1 Avg	F1 Prop*
Bias Bluff Busters	4	2	8	2,600	31.5	54.3	47.3	62.6	20.2	51.1	**37.6**	54.2
Narrative Navigators	7	2	4	2,200	12.8	54.5	54.6	63.1	19.4	52.7	37.1	60.7
NLPColab	21	15	30	**16,500**	**69.9**	87.3	73.1	92.1	21.5	53.3	**39.9**	61.6
Sahara Pioneers	4	2	8	2,600	27.9	49.1	46.3	50.5	18.7	48.7	**37.4**	56.3
Sina	10	12	24	**13,200**	**65.3**	85.7	67.9	76.7	12.5	48.1	27.7	44.5
The CyberEquity Lab	5	3	15	**4,200**	**33.5**	65.3	41.1	54.2	22.1	55.0	**37.4**	55.7
Average	8.5	6.0	14.8	6,883	40.1	66.0	55.0	66.5	19.1	51.5	36.2	55.5
Total	51	36	89	41,300								

Table 4: Results of the Propaganda subtask.

6 Discussion

We now consider the label distribution patterns for Bias and Propaganda, independently and together.

6.1 Bias Label Distributions

Table 6 summarizes the Bias label distributions overall and for different source languages: Arabic (Ar), English (En), French (Fr), Hebrew (He), and Hindi (Hi). The reported results here include all annotated data points (from Main and IAA). Naturally, data points from Batches 1 and 2 are over-represented since they were annotated by all teams. **Unbiased** was the highest overall label claiming over two-fifths of all data points. This is followed by **Biased against Palestine** (29.2%), then **Biased against Israel**; with **Biased against Palestine** almost three times that of **Biased against Israel**. **Biased against others** appeared in about 1/16th of the cases. All labels seem to be generally equally distributed across source languages with the exception of **Biased against Palestine** and **Biased against Israel** standing out in Hebrew-sourced texts: **Biased against Palestine** is twice the average of other languages; and **Biased against Israel** is one-fifth the average of other languages. While this seems consistent with what would be expected of the news media of a country at war, it is possible that there is an additional priming bias from knowing the source language of the text. Perhaps in future editions, we could compare with a setting where all source language information is hidden and only translations are provided.

Subtask	Track	1st Place	2nd Place	3rd Place
Bias	**Guidelines**	NLPColab	Eagles	Narrative Navigators
Bias	**IAA Quality**	NLPColab	JusticeLeague	Sina
Bias	**Quantity**	DRAGON	NLPColab	Sina
Bias	**Consistency**	The Lexicon Ladies	NLPColab	Narrative Navigators
Propaganda	**Guidelines**	NLPColab	Bias Bluff Busters	Sina
Propaganda	**IAA Quality**	NLPColab	Sina	The CyberEquity Lab
Propaganda	**Quantity**	NLPColab	Sina	The CyberEquity Lab
Propaganda	**Consistency**	NLPColab	Bias Bluff Busters	Sahara Pioneers/The CyberEquity Lab

Table 5: Shared task winners for each subtask and track.

	Source Language					
Label	**Arabic**	**English**	**French**	**Hebrew**	**Hindi**	**Total**
Unbiased	9.5%	9.7%	8.6%	6.0%	9.0%	**42.8%**
Biased against Palestine	5.3%	4.6%	4.1%	**10.4%**	4.7%	**29.2%**
Biased against Israel	2.7%	2.2%	3.5%	**0.5%**	2.0%	**10.9%**
Biased against others	1.1%	1.4%	1.3%	1.1%	1.4%	6.4%
Unclear	1.0%	1.0%	1.4%	1.3%	1.7%	6.4%
Not Applicable	0.3%	0.7%	0.7%	0.6%	0.6%	2.8%
Biased against both Palestine and Israel	0.2%	0.4%	0.3%	0.1%	0.6%	1.6%
Total	20.0%	20.0%	20.0%	20.0%	20.0%	100.0%

Table 6: Bias Label Distributions in total and over source language.

	Source Language					
Label	**Arabic**	**English**	**French**	**Hebrew**	**Hindi**	**Total**
Propaganda	9.2%	8.3%	8.5%	**11.7%**	8.4%	**46.1%**
Not Propaganda	9.6%	10.1%	9.2%	**6.6%**	9.8%	**45.2%**
Unclear	0.9%	0.9%	1.6%	1.1%	1.4%	6.0%
Not Applicable	0.3%	0.7%	0.7%	0.5%	0.4%	2.7%
Total	20.0%	20.0%	20.0%	20.0%	20.0%	100.0%

Table 7: Propaganda Label Distributions in total and over source language.

	Propaganda Labels			
Bias Labels	**Propaganda**	**Not Propaganda**	**Not Applicable**	**Unclear**
Biased against Palestine	**74.3%**	-61.2%	-20.1%	-21.7%
Biased against Israel	-7.2%	**15.2%**	-12.5%	-6.4%
Biased against both Palestine and Israel	-9.0%	**13.7%**	-7.2%	-3.2%
Biased against others	**23.1%**	-20.0%	-7.0%	-4.0%
Unbiased	-69.8%	**78.7%**	-10.4%	-2.3%
Not Applicable	-27.9%	-14.7%	**85.9%**	25.5%
Unclear	-25.9%	-9.7%	31.2%	**57.6%**

Table 8: Pearson correlation coefficient for all Bias vs Propaganda label pairs.

6.2 Propaganda Label Distributions

Table 7 summarizes the Propaganda label distributions overall and for different source languages. The reported results here include all annotated data points (from Main and IAA). **Propaganda** and **Not Propaganda** split the distribution almost equally with **Propaganda** being slightly higher. All labels seem to be generally equally distributed with the exception of of Hebrew-sourced texts where propaganda spikes: the ratio of **Propaganda** to **Not Propaganda** is 1.8 as opposed to less than 0.9 for the other source languages. The same observation made in the previous section applies here.

Lang	Text	English	Unbiased	Biased > Palestine	Biased > Israel	Propaganda	Not Propaganda
Hi	इसराइल-हमास के बीच अस्थाई युद्धविराम, वापस लौट रहे लोगों ने क्या कहा	Temporary ceasefire between Israel and Hamas, what the returning people said	**94%**				**100%**
En	Pro-Palestinian protesters marched in New York City on Sunday to call for a ceasefire between Israel and Hamas amid New Year's Eve celebrations.		**94%**				**100%**
Ar	فرحة كبيرة بالقدس والضفة الغربية بعد الإفراج عن أسرى المرحلة الأولى من صفقة التبادل بين حماس وإسرائيل.	Great joy in Jerusalem and the West Bank after the release of prisoners from the first phase of the exchange deal between Hamas and Israel.	**88%**				**100%**
He	נזכור ולא נשכח את הנרצחים והנופלים על הגנת המולדת. נלחם עד שהנאצים של חמאס יושמדו. עם ישראל חי! 💙 🇮🇱	We will remember and not forget those who were murdered and who fell in defense of the homeland. Fight until the Hamas Nazis are destroyed. Israel Lives! 💙 🇮🇱		**88%**		**100%**	
Fr	Le Hamas se cache délibérément parmi les civils, faisant ainsi payer aux Gazaouis les conséquences des atrocités commises par le Hamas. Notre guerre est contre le Hamas, et non contre la population de Gaza. Nous prenons des mesures importantes pour minimiser les dommages causés aux civils, alors que le Hamas les utilise comme boucliers humains.	Hamas deliberately hides among civilians, making Gazans pay for the consequences of Hamas's atrocities. Our war is against Hamas, not against the people of Gaza. We are taking important steps to minimize harm to civilians while Hamas uses them as human shields.		**88%**		**100%**	
Ar	حماس فتحت أبواب الجحيم على قطاع غزة #السيوف_الحديدية	Hamas opened the gates of hell on the Gaza Strip #Iron_Swords		**88%**		**100%**	
Fr	Cette nuit à Gaza...le massacre des civils se poursuit #GazaUnderAttack #Palestine	Tonight in Gaza...the massacre of civilians continues #GazaUnderAttack #Palestine	31%	13%	**56%**	**83%**	17%
En	Hamas has invited Elon Musk to witness in person the scope of the violence and devastation heaped upon the Gaza Strip by Israel		31%		**63%**	**50%**	**50%**
Hi	इसराइल के हमलों से अस्पताल भी अछूते नहीं... युद्ध-विराम ख़त्म होने के बाद से इसराइल और हमास के बीच एक बार फिर युद्ध छिड़ चुका है. इसराइल अब दक्षिणी ग़ज़ा को भी निशाना बना रहा है, जिससे अस्पताल भी अछूते नहीं हैं.	Even hospitals are not untouched by Israeli attacks… Since the end of the ceasefire, war has once again broken out between Israel and Hamas. Israel is now targeting Southern Gaza as well, from which even hospitals are not untouched.	25%	13%	**56%**	**50%**	**50%**

Table 9: Examples of different texts and their most relevant annotation distributions.

6.3 Bias vs Propaganda Label Correlations

Finally, Table 8 presents a cross-comparison of Bias and Propaganda labels using Main Batches 1 and 2 ($1,800$ data points). For each text in this data subset, we calculate for each label in Bias and Propaganda subtasks the number of teams that selected that label. We then report the Pearson product-moment correlation coefficient of the $1,800$ counts for every pair of Bias-Propaganda labels. Unsurprisingly, Bias **Not Applicable** and Propaganda **Not Applicable** relate closely (r=85.9%). Similarly r=78.7% for **Unbiased** is **Not Propaganda**. The **Biased against Palestine** label correlates positively highly (74.3%) with **Propaganda**, and negatively strongly (-61.2%) with **Not Propaganda**. The patterns are reversed and much weaker for **Biased against Israel**. These are only high level initial observations that open exciting possibilities for further study.

6.4 Selected Examples

Table 9 presents a number of texts with their associated most prominent label averages across the participating teams. The examples are from Main Batches 1 and 2. There are nine examples; the first three were marked by the vast majority of teams as **Unbiased** and **Not Propaganda**. The second set of three examples were marked by the vast majority again as **Biased against Palestine** and **Propaganda**. The last set shows examples with majority **Biased against Israel**.

7 Conclusion and Future Work

The FIGNEWS shared task successfully brought together a diverse community to annotate bias and propaganda in multilingual news posts. This initiative brought together 17 teams producing 129,800 data points. The shared task highlighted the crucial role of clear guidelines, examples, and collaboration in advancing NLP research on complex, subjective, and sensitive, opinion analysis tasks. The resulting dataset and insights contribute valuable resources and direction for future work in this important area. All data and code are publicly available.[3]

Future work should focus on expanding the annotation efforts to include more diverse languages and topics, and refining annotation guidelines based on participant feedback. The created data should be leveraged to advance NLP automatic bias and propaganda detection techniques, as well as foster interdisciplinary studies to deepen our understanding of bias and propaganda in news media.

Limitations

We acknowledge the following limitations in the FIGNEWS shared task design and implementation.

- **Annotation Subjectivity** The task involves subjective judgments on bias and propaganda, which vary among annotators and teams, impacting annotation consistency and reliability. Teams' self-selection based on preconceived notions about the topic may further influence this variability.
- **Label Selection** We acknowledge that the set of labels we specified limit the space of possibilities and may oversimplify complex issues imposing a binary perspective that does not fully capture nuanced viewpoints and biases within the dataset.
- **Scope of Topics and Text Selection** The focus on the early days of the Israel War on Gaza may limit the applicability of findings to other types of news events or broader media contexts. The size of the corpus is relatively small, and may include some sampling bias.
- **Limited Diversity** while we observed a range of backgrounds among the annotators, we acknowledge that some groups were highly over-represented, which potentially biases the overall conclusions.

Ethical Considerations

The FIGNEWS shared task deals with sensitive topics and media narratives related to the Israel War on Gaza. The organizers and participants have taken several measures to ensure ethical considerations are addressed:

- **Anonymization** All posts have been anonymized, with no identifying information about the account owners or users provided to the participants.
- **Public Posts** Only publicly available posts from verified accounts were included in the dataset, ensuring that the content was intended for public dissemination.
- **Balanced Representation** To ensure fair representation, the dataset includes a balanced number of posts from various viewpoints and narratives during the war.
- **Responsible Use** Participants were required to agree to use the dataset solely for research purposes and not for any unethical or illegal activities.

Acknowledgments

This task is partially supported by NPRP 14C-0916-210015 from the Qatar National Research Fund, which is a part of Qatar Research Development and Innovation Council (QRDI). The findings achieved herein are solely the responsibility of the authors.

References

Dmitrii Aksenov, Peter Bourgonje, Karolina Zaczynska, Malte Ostendorff, Julián Moreno-Schneider, and Georg Rehm. 2021. Fine-grained classification of political bias in german news: A data set and initial experiments. In *WOAH 2021 - 5th Workshop on Online Abuse and Harms, Proceedings of the Workshop*.

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.

Maryam Al Emadi, Jana El Messelmani, Lyna Bermak, Goumana Abdullah, Esraa Sherqawi, Anissa Jrad, Zied Zouabi, and Wajdi Zaghouani. 2024. Narrative Navigators at FIGNEWS 2024 shared task: New frontiers in bias and propaganda annotation techniques. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok. Association for Computational Linguistics.

Asmahan Juma Al-Mamari and Fatma Mohammed Al-Farsi. 2024. SQUad at FIGNEWS 2024 shared task: Unmasking bias in social media through data analysis and annotation. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. 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. In *Proceedings of the International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, pages 98–103.

Al Manar Al Wardi, Blqees Al Busaidi, Malath Al Sibani, Hiba Al Siyabi, and Najma Al Zidjaly. 2024. BiasGanda at FIGNEWS 2024 shared task: A quest to uncover biased views in news coverage. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok. Association for Computational Linguistics.

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. In *Journal of Economic Perspectives*, volume 31, pages 211–236.

Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2020. What was written vs. who read it: News media profiling using text analysis and social media context. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3363–3374.

Alberto Barrón-Cedeno, Israa Jaradat, Giovanni Da San Martino, and Preslav Nakov. 2019. Proppy: Organizing the news based on their propagandistic content. *Information Processing & Management*, 56(5):1849–1864.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 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.

Biasly. 2017. Biasly. https://www.biasly.com/politician-ratings/.

Ghizlane Bourahouat and Samar M. Amer. 2024. The Guidelines Specialists at FIGNEWS 2024 shared task: An annotation guideline to unravel bias in news media narratives using a linguistic approach. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. 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(2):250–271.

Marco Casavantes, Manuel Montes y Gómez, Luis Carlos González, and Alberto Barróñ-Cedeno. 2024. Propitter: A twitter corpus for computational propaganda detection. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14392 LNAI.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Kareem Darwish, Nizar Habash, Mourad Abbas, Hend Al-Khalifa, Huseein T. Al-Natsheh, Houda Bouamor, Karim Bouzoubaa, Violetta Cavalli-Sforza, Samhaa R. El-Beltagy, Wassim El-Hajj, Mustafa Jarrar, and Hamdy Mubarak. 2021. A Panoramic survey of Natural Language Processing in the Arab Worlds. *Commun. ACM*, 64(4):72–81.

Lina Duaibes, Areej Jaber, Mustafa Jarrar, Ahmad Qadi, and Mais Qandeel. 2024. Sina at FIGNEWS 2024: Multilingual datasets annotated with bias and propaganda. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok. Association for Computational Linguistics.

Amanda C. Ean, Mai A. Baddar, and Sofien Bâazaoui. 2024. Eagles at FIGNEWS 2024 shared task: A context-informed prescriptive approach to bias detection in contentious news narratives. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Yousra Alaa El-Ghawi, Abeer Ehab Marzouk, and Aya Mohamed Khamis. 2024. Lexicon Ladies at FIGNEWS 2024 shared task: Identifying keywords for bias annotation guidelines of facebook news headlines on the Israel-Palestine 2023 war. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

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, Pratthana 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.

Nina Gorenc. 2020. Political communication in post-truth society: The case of the 2016 us election. *Ars et Humanitas*, 14.

Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378.

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

Nagham Hamad, Mustafa Jarrar, Mohammed Khalilia, and Nadim Nashif. 2023. Offensive Hebrew Corpus and Detection using BERT. In *Proceedings of the 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*. IEEE.

Felix Hamborg, Anastasia Zhukova, and Bela Gipp. 2019. Automated identification of media bias by word choice and labeling in news articles. In *Proceedings of the 19th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, pages 1–10.

Mohammed H.S. Helal, Radi Jarrar, Mohammad Alkhanafseh, Abdallah Karakra, and Ruba Awadallah. 2024. The CyberEquity Lab at FIGNEWS 2024 shared task: Annotating a corpus of facebook posts to label bias and propaganda in gaza-Israel war coverage in five languages. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Freddy Heppell, Kalina Bontcheva, and Carolina Scarton. 2023. Analysing state-backed propaganda websites: a new dataset and linguistic study. In *EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings*.

Jiwoo Hong, Yejin Cho, Jaemin Jung, Jiyoung Han, and James Thorne. 2023. Disentangling structure and style: Political bias detection in news by inducing document hierarchy. In *Findings of the Association for Computational Linguistics: EMNLP 2023*.

Benjamin D Horne and Sibel Adali. 2017. Just how toxic is data? assessing the impact of data sources on fake news detection. In *Proceedings of the 10th International Conference on Web and Social Media (ICWSM)*, pages 259–268.

Sadegh Jafari, Mohsen Mahmoodzadeh, Vanooshe nazari, Razieh Bahmanyar, and Kathryn Burrows. 2024. DRAGON at FIGNEWS 2024 shared task: A dedicated rag for october 7th conflict news. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Mustafa Jarrar, Nagham Hamad, Mohammed Khalilia, Bashar Talafha, AbdelRahim Elmadany, and Muhammad Abdul-Mageed. 2024. WojoodNER 2024: The Second Arabic Named Entity Recognition Shared Task. In *Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Lalitha Kameswari, Dama Sravani, and Radhika Mamidi. 2020. Enhancing bias detection in political news using pragmatic presupposition. In *Proceedings of the Eighth International Workshop on Natural Language Processing for Social Media*, pages 1–6, Online. Association for Computational Linguistics.

Khalid Al Khatib, Sara Gemelli, Saskia J. Heisterborg, Gosse Minnema Pritha Majumdar, Arianna Muti, and Noa Visser Solissa. 2024. Groningen Annotates Gaza at the FIGNEWS 2024 shared task: Analyzing bias in conflict narratives. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Kang Min Kim, Mingyu Lee, Hyun Sik Won, Min Ji Kim, Yeachan Kim, and Sang Keun Lee. 2023. Multi-stage prompt tuning for political perspective detection in low-resource settings. *Applied Sciences (Switzerland)*, 13.

David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. In *Science*, volume 359, pages 1094–1096.

Xiaodong Li, Lidong Bing, Wai Lam, and Bei Shi. 2019. A survey of sentiment analysis in social media. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*, pages 3127–3137.

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 *Proceedings of the 2018 ACM on Conference on Information and Knowledge Management (CIKM)*, pages 2482–2484.

Iffat Maab, Edison Marrese-Taylor, and Yutaka Matsuo. 2023. Target-aware contextual political bias detection in news. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 782–792, Nusa Dua, Bali. Association for Computational Linguistics.

Giovanni Da San Martino, Alberto Barrón-Cedeno, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020a. Semeval-2020 task 11: Detection of propaganda techniques in news articles. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1377–1414.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020b. A survey on computational propaganda detection. *IJCAI International Joint Conference on Artificial Intelligence*, 2021-January.

Jacinta Mwende Maweu. 2019. "fake elections"? cyber propaganda, disinformation and the 2017 general elections in kenya. *African Journalism Studies*, 40.

Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen Corman, and Huan Liu. 2018. Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1(2):1–20.

Abdusalam F. A. Nwesri, Mai Elbaabaa, Fatima Benlashihar, and Fatma Alalos. 2024. UoT1 at FIGNEWS 2024 shared task: Labeling news bias. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok. Association for Computational Linguistics.

Silvia Pareti, Jasmin Heierli, Serena Pareti, and Tatiana Lando. 2024. Bias Bluff Busters at FIGNEWS 2024 shared task: Developing guidelines to make bias conscious. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. Newscube: Delivering multiple aspects of news to mitigate media bias. In *Proceedings of the 27th International Conference on*

Human Factors in Computing Systems (CHI), pages 443–452.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937.

Sadaf Abdul Rauf, Saadia Ishtiaq Nauman, Huda Sarfraz, Hammad Afzal, Arooj Fatima, Sadaf Ziafat, Momina Ishfaq, and Alishba Abdul Suboor. 2024. NLPColab at FIGNEWS 2024 shared task: Bias and propaganda identification. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Francisco Javier Rodrigo-Ginés, Jorge Carrillo de Albornoz, and Laura Plaza. 2024. A systematic review on media bias detection: What is media bias, how it is expressed, and how to detect it.

Valle Ruiz-Fernández, José Javier Saiz, and Aitor Gonzalez-Agirre. 2024. BSC-LANGTECH at FIGNEWS 2024 shared task: Exploring semi-automatic bias annotation using frame analysis. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Noor Sadiah, Sumaya Abdul Rahman, and Sara Al-Emadi. 2024. Ceasefire at FIGNEWS 2024 shared task: Automated detection and annotation of media bias using large language models. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Amr Khaled Saleh, Huda Maher Mohamed, and Hager Sayed Rashed. 2024. JusticeLeague at FIGNEWS 2024 shared task: Innovations in bias annotation. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Karishma Sharma, Feng Qian, Haifeng Jiang, Natali Ruchansky, Minghui Zhang, and Yan Liu. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3):1–42.

Manan Sharma, Krishanu Kashyap, Kushagra Gupta, and Shailender Kumar. 2023. Advancing political bias detection: A novel high-accuracy model. In *Proceedings of International Conference on Computational Intelligence and Sustainable Engineering Solution, CISES 2023*.

Marwa Solla, Hassan Ebrahem, Alya Issa, Harmain Harmain, and Abdusalam Nwesri. 2024. Sahara Pioneers at FIGNEWS 2024 shared task: Data annotation guidelines for propaganda detection in news items. In *The Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.

Veronika Solopova, Viktoriia Herman, Christoph Benzmüller, and Tim Landgraf. 2024. Check news in one click: Nlp-empowered pro-kremlin propaganda detection. In *EACL 2024 - 18th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of System Demonstrations*.

Timo Spinde, Jan-David Krieger, Manuel Plank, and Bela Gipp. 2021a. Towards a reliable ground-truth for biased language detection. *arXiv preprint arXiv:2112.07421*.

Timo Spinde, Kanishka Sinha, Norman Meuschke, and Bela Gipp. 2021b. Tassy: A text annotation survey system. *arXiv preprint arXiv: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):1–23.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

Anuradha Yenkikar, C. Narendra Babu, and D. Jude Hemanth. 2022. Sentinet: a deep sentiment analysis network for political media bias detection. *Dyna (Spain)*, 97.

Wajdi Zaghouani, Eman Awad, and Anis Charfi. 2016. Error annotation guidelines for arabic texts. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 3483–3488.

Wajdi Zaghouani and Anis Charfi. 2018. Annotation guidelines for text analytics in social media. In *Qatar Foundation Annual Research Conference Proceedings 2018*, page SSHARP1191.

Xiang Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.

Yang Zhou, Paul Resnick, and Qiaozhu Mei. 2020. Detecting political bias in social media content. In *Proceedings of the 14th International AAAI Conference on Web and Social Media (ICWSM)*, pages 299–308.

A Shared Task Details

We provide below an updated version of the shared task details reflecting the final decisions made in the effort, e.g., we added a fourth evaluation track (IAA Quality) that was not mentioned in the original call to participate.

A.1 Shared Task Objectives

The shared task aims to serve as a collaborative platform where participants propose guidelines and diverse methods for annotating and analyzing the dataset.

Provided Data The organizers will provide 15 batches of social media posts, 1,000 post per batch. Each 1,000-post batch will contain 200 posts from 5 languages: Arabic, English, French, Hebrew, and Hindi, together with their machine translated versions into Arabic and English (as needed). Participants must specify whether they annotated the original language or its machine translated version. So a monolingual Arabic team can annotate the full batch in Arabic (original or translated). The batches will be provided to the annotation teams with clear instruction on how to submit the results.

Minimal Annotations to Qualify To qualify, each participating team **must provide full annotation guidelines** for each subtask they choose to work on; and they **must annotate a minimum of two batches**, i.e. (1,800 posts) and their designated Inter-annotator agreement subset (200 posts) for a total of 2,000 posts (specifically Batch 1 and Batch 2). The IAA subset must be done by every annotator on the team; but the rest can be divided among them.

A.2 Shared Task Subtasks and Tracks

There will be two **subtasks** of focus. For each subtask, there will be four **evaluation tracks** for which winners will be crowned.

A.2.1 Subtask on Bias Annotation

The subtask is restricted to seven possible Labels, presented below with illustrative examples.

1. **Unbiased**
   Example: "In the ongoing Israel-Palestine conflict, recent events have escalated tensions. Yesterday, Israeli forces conducted operations in response to rocket attacks from Gaza. Both sides have reported casualties. International leaders are calling for restraint and a return to peace talks."
2. **Biased against Palestine**
   Example: "Once again, Palestinian aggression has disrupted peace in the region. Palestinian extremists, ignoring efforts for peace, launched unprovoked attacks on innocent Israeli civilians. Israel's response, though portrayed as harsh by some, is a justified measure to protect its citizens."
3. **Biased against Israel**
   Example: "In a typical display of excessive force, Israeli troops have yet again targeted Palestinian areas, causing numerous civilian casualties. This aggression, under the guise of self-defense, highlights the ongoing oppressive tactics Israel employs against Palestinians."
4. **Biased against both Palestine and Israel**
   Example: "In the latest chapter of their endless and futile conflict, Israeli and Palestinian forces have once again engaged in senseless violence. Both sides continue to commit atrocities, showing a complete disregard for peace or human life."
5. **Biased against others**
   Example: "In the shadow of the Israel-Palestine conflict, external actors, particularly Iran, are exacerbating tensions. Iran's covert support for extremist groups shows its intent to destabilize the region, disregarding the catastrophic impact on both Israeli and Palestinian civilians."
6. **Unclear**
   Example: "Recent developments in the Middle East have seen an increase in hostilities. The situation in the region is complex, with various factors contributing to the current state of affairs. The international community remains divided on the issue."

7. **Not Applicable**
Example: "In other news, the annual technology conference in Tel Aviv has unveiled groundbreaking advancements in cybersecurity. Industry leaders from around the globe gathered to showcase innovations that promise to shape the future of digital security."

A.2.2 Subtask on Propaganda Annotation

The subtask is restricted to four possible Labels, presented below with illustrative examples.

1. **Propaganda**
Example: "In a display of unmatched heroism, our troops have once again safeguarded our nation from the brink of destruction, heroically neutralizing the threat from Gaza, which aims to undermine our very existence."
2. **Not Propaganda**
Example: "Yesterday, an escalation occurred along the Israel-Gaza border, resulting in casualties on both sides. Israeli and Palestinian officials provided conflicting accounts of the events that led to the confrontation."
3. **Unclear**
Example: "The situation in Gaza remains tense, with reports of civilian distress and military movements. While some sources claim the military actions are defensive, others argue they are provocative, leaving the true nature of the situation open to interpretation."
4. **Not Applicable**
Example: "A feature on Gaza's cultural scene highlights the resilience of its art community, showcasing how local artists use their craft to express hope and endurance amid challenging circumstances, without delving into the political context."

A.2.3 Guidelines Evaluation Track

The teams have the freedom to design their own guidelines and apply them to the shared data. The following is the checklist of all items that will be evaluated by the organizers.

Annotation Guidelines Detailed annotation guidelines including examples for all main and corner cases. Consider the following components which will be used in the evaluation of the guidelines.

1. **Define the Objective and Describe the Task**
Outline the purpose and specific NLP task. Provide a detailed task description with correct examples.
2. **Establish Categories**
List and define all annotation labels/categories/tags.
3. **Include Detailed Category Guidelines with Examples**
Explain application criteria for each category/tag, with examples. Offer examples for correct application and common mistakes.
4. **Outline the Process**
Describe the step-by-step annotation process and tools used.
5. **Set Quality Standards**
Define expectations for accuracy and consistency, along with quality check procedures.
6. **Handle Ambiguities and Difficult Cases**
Provide guidance on ambiguous cases and a protocol for seeking clarification.
7. **Ensure Consistency**
Implement measures for annotator consistency and recommend calibration sessions.
8. **Training and Support**
Detail training procedures and support resources for annotators. Highlight unbiased annotation practices and handling of sensitive data. Schedule guideline reviews for updates based on feedback and new insights. Include a system for annotator feedback to refine guidelines and processes.

The 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 Guidelines Score used to determine the winners of this track is the average normalized Document Score and IAA Kappa score.

$$\text{GuidelinesScore}_i = \text{Average}\left(\frac{\text{DocumentScore}_i}{\text{DocumentScore}_{max}}, \frac{\text{IAAKappa}_i}{\text{IAAKappa}_{max}}\right)$$

The Document Score is equal to the number of satisfied document components mentioned above, a range from 0 to 8. The IAA Kappa score of a team is the average of all pairwise IAA Kappas over team annotators per batch.

A.2.4 IAA Quality Evaluation Track

We will also report the IAA Kappa scores per team and use them to determine the best performers, independent of the guidelines track.

A.2.5 Quantity Evaluation Track

The teams can compete in the number of annotated data batches. They must finish them in order and complete a batch before moving to the next. The teams with the highest number of completed batches will be crowned the Quantity track winners for the subtask of choice.

A.2.6 Consistency Evaluation Track

The various teams in the same subtask and shared completed batches will be compared for correlation against each other. The teams that have the highest correlation against other teams (centroidal choices) will be crowned winner. This needs a minimum of three teams per subtask.

A.3 Publication

All teams participating in the shared task are invited to submit short paper (4 pages) descriptions of their efforts. These papers will be evaluated by multiple reviewers to be selected for publication in the ArabicNLP 2024 Conference Proceedings and indexed by the ACL Anthology.

A.4 Collaborative Commitment

Participants are encouraged to join the shared task with a commitment to collaboration. Whether working independently or within teams, every effort and insight contributed should be shared openly. This collaborative ethos extends beyond individual tasks and includes sharing methodologies, findings, and results.

A.5 Optional Demographic Details

We would like to invite participants to provide some demographic details voluntarily. This information includes aspects such as age range, native language, educational background, area of study or expertise, gender, and region of origin. Please note that providing this demographic information is entirely optional and will not influence the evaluation of your participation in any way. We respect your privacy and understand if you choose not to share these details.

B Annotation Interface

	A	B	C	D	E	F	G	H	I	J	K
1	Batch	Source Language	ID	Type	Text	English MT	Arabic MT	Annotator ID	Bias	Propaganda	Comments
2	B01	English	1	MAIN	Yemen's Houthis have wad	Yemen's Houthis have wad	خاض الحوثيون في اليمن الحرب بين				
3	B01	Arabic	1	MAIN	الارهابية الداعشية تجلب العار للعرب	Hamas, the ISIS terrorist or	حماس الارهابية الداعشية تجلب العار				
4	B01	French	1	MAIN	De la nourriture pour anima	Animal food transformed in	تحويل الغذاء الحيواني إلى دقيق ومن				
5	B01	Hebrew	1	MAIN	the whole world can see it!	On the one hand, what hap	من ناحية، ماذا يحدث عندما يلتقي جنو				
6	B01	Hindi	1	MAIN	LIVE- अमेरिका-इजराइल मिलक	LIVE- America-Israel will to	بث مباشر- أمريكا وإسرائيل ستهاجمان				
7	B01	English	2	MAIN	Isreal - Hamas Conflict | Fa	Isreal - Hamas Conflict | Fa	إسرائيل - الصراع مع حماس | وجها				
8	B01	Arabic	2	MAIN	المكتب السياسي لحركة "حماس" إسم	Head of the Hamas Politica	رئيس المكتب السياسي لحركة "حماس"				
9	B01	French	2	MAIN	Samedi 7 octobre, il était au	On Saturday October 7, it w	في يوم السبت 7 أكتوبر، كانت الساعة				
10	B01	Hebrew	2	MAIN	world would react to this!!!	On the morning of a Jewish	في صباح يوم عطلة يهودية في إسرائيل				
11	B01	Hindi	2	MAIN	LIVE | हमास पर 180 डिग्री वाले	LIVE | 180 degree attacks	مباشر | تبدأ الهجمات 180 درجة على				
12	B01	Arabic	3	MAIN	يصفه الإعلام الغربي بأنه "مرعب إسرا	The Western media describ	يصفه الإعلام الغربي بأنه "مرعب إسر				
13	B01	French	3	MAIN	39 Gazaouis blessés, dont	39 injured Gazans, includin	تم علاج 39 مصابًا من سكان غزة، بيا				
14	B01	Hebrew	3	MAIN	Gaza. My interview on RT.	I will not allow any presente	لن أسمح لأي مذيع يجري مقابلتي في				
15	B01	English	4	MAIN	Videos show how armed m	Videos show how armed m	أظهرت مقاطع فيديو كيف اقتحم مسلحو				
16	B01	Arabic	4	MAIN	أطلقنا #السيوف_الحديدية ضد حماس و	We fired #iron_swords agai	أطلقنا #السيوف_الحديدية ضد حماس				
17	B01	Hebrew	4	MAIN	זאת סבתא שלי! נ. Adva Adar	This is my grandmother! Yo	هذه جدتي! لقد تم اختطافك دون عوائق				
18	B01	Hindi	4	MAIN	अगर इजरायल और ईरान में हुई ज	If there is a war between Is	إذا اندلعت حرب بين إسرائيل وإيران				
19	B01	French	5	MAIN	Gaza était la plus grande pr	Gaza was the largest open-	كانت غزة أكبر سجن مفتوح في العالم.				
20	B01	Hebrew	5	MAIN	ation must be made extinct.	Hamas leader Ismail Haniy	زعيم حماس إسماعيل هنية يحتفل من				
21	B01	Hindi	5	MAIN	हमास और इजरायल के युद्ध के क	Many days have passed sir	لقد مرت أيام عديدة على الحرب بين				
22	B01	French	6	MAIN	🚨 Un carnage humanitaire	🚨 Humanitarian carnage is	🚨المذبحة الإنسانية مستمرة في غزة.				
23	B01	Hindi	6	MAIN	इजरायल की सेना से हमास के आ	What did the Hamas terrori	ماذا قال إرهابي حماس للجيش الإسرائ				
24	B01	English	7	MAIN	Protest in Aligarh Muslim U	Protest in Aligarh Muslim U	وقفة احتجاجية في جامعة عليكرة الإسلا				
25	B01	Arabic	7	MAIN	سة لـ #الحدث مع المتحدث باسم الجيش	A special interview for the #	مقابلة خاصة لـ #الحدث مع المتحدث ب				
26	B01	French	7	MAIN	Vous en pensez quoi des tw	What do you think of Benze	ما رأيكم بتغريدات بنزيما يا أحباب؟				
27	B01	Hindi	7	MAIN	👉 हमास की हिमाकत..मोसाद की	👉 Hamas' audacity..Mossa	👉 جرأة حماس.. فشل الموساد! 👉				
28	B01	English	8	MAIN	IDF releases audio recordir	IDF releases audio recordir	الجيش الإسرائيلي ينشر تسجيلًا صوتيًا				
29	B01	Arabic	8	MAIN	قق أي إنجاز ميداني على الأرض سوى	"The enemy did not achieve	"العدو لم يحقق أي إنجاز ميداني على				
30	B01	Hebrew	8	MAIN	לוחמינו היקרים ערכו אמש את ו	Our dear soldiers last night	جنودنا الأعزاء أقاموا ليلة أمس صلاة				
31	B01	Hindi	8	MAIN	इज़राइल-हमास युद्ध के बहाने ट्विट	There is a flood of fake new	هناك طوفان من الأخبار الكاذبة على تو				
32	B01	French	9	MAIN	Malgré la trêve entrée en vi	Despite the truce that came	رغم الهدنة التي دخلت حيز التنفيذ يوم				
33	B01	Hebrew	9	MAIN	🇮🇱 أهلا حماس، انا كثير مبسوط اعر	Hello Hamas, I am so happ	مرحباً حماس، أنا سعيد جدًا بمعرفة أنك				
34	B01	Hindi	9	MAIN	#LIVE : सुरंगो में छिपे हमास के	#LIVE: Hamas fighters hidir	#مباشر: مقاتلو حماس يختبئون في الأ				
35	B01	English	10	MAIN	CNN's Dana Bash pressed	CNN's Dana Bash pressed	حث دانا باش، مراسل سي إن إن، رئيه				
36	B01	French	10	MAIN	Tom Sisley, franco-israélien	Tom Sisley, Franco-Israeli,	توم سيسلي، الفرنسي الإسرائيلي، "يبكو				
37	B01	Hindi	10	MAIN	पाकिस्तानी टीम क्रिकेट खेलने आई	Pakistani team came to pla	المنتخب الباكستاني جاء للعب الكريكيت				
38	B01	Arabic	11	MAIN	أسيرة إسرائيلية مُفرج عنها: مددت ي	A released Israeli prisoner:	أسيرة إسرائيلية مُفرج عنها: مددت يدي				
39	B01	French	11	MAIN	« Vous voulez qu'on se batt	"You want us to fight all our	"هل تريد منا أن نقاتل طوال حياتنا؟ أ				
40	B01	Hebrew	11	MAIN	is launched a war on Israel!	Hamas has declared war o	حماس أعلنت الحرب على إسرائيل! ش				
41	B01	Hindi	11	MAIN	#IsraelPalestineWar: इजराय	#IsraelPalestineWar: How l	#الحرب_الإسرائيلية_الفلسطينية: إلى				

Figure 1: A screenshot of the Google Sheet annotation setup for the Main data subset.

C Annotator Demographics

Native Language	Bias	Propaganda	Total	Total (%)
Arabic	43	24	67	49.3%
Urdu	21	21	42	30.9%
Italian	4	2	6	4.4%
Persian	4	0	4	2.9%
Dutch	3	0	3	2.2%
English	2	1	3	2.2%
German	1	1	2	1.5%
Hindi/Urdu	2	0	2	1.5%
Russian	1	1	2	1.5%
Spanish	2	0	2	1.5%
Bengali	1	0	1	0.7%
Prefer not to say	1	1	2	1.5%
Total	85	51	136	100.0%

Age Range	Bias	Propaganda	Total	Total (%)
18-24	48	17	65	47.8%
25-34	20	19	39	28.7%
35-44	14	9	23	16.9%
45-54	1	1	2	1.5%
55-64	1	0	1	0.7%
Prefer not to say	1	5	6	4.4%
Total	85	51	136	100.0%

Gender	Bias	Propaganda	Total	Total (%)
Female	67	38	105	77.2%
Male	16	8	24	17.6%
Non-binary	1	0	1	0.7%
Prefer not to say	1	5	6	4.4%
Total	85	51	136	100.0%

Region of Origin	Bias	Propaganda	Total	Total (%)
South Asia	24	21	45	33.1%
Levant	19	17	36	26.5%
North Africa	16	2	18	13.2%
Western Europe	11	4	15	11.0%
GCC Countries	7	0	7	5.1%
Other Middle Eastern Countries	4	0	4	2.9%
Southeast Asia	1	0	1	0.7%
Prefer not to say	3	7	10	7.4%
Total	85	51	136	100.0%

Education Level	Bias	Propaganda	Total	Total (%)
Master's degree	36	27	63	46.3%
Bachelor's degree	38	12	50	36.8%
Doctoral degree	8	6	14	10.3%
Post-doctoral training	1	0	1	0.7%
Prefer not to say	2	6	8	5.9%
Total	85	51	136	100.0%

Area of Expertise	Bias	Propaganda	Total	Total (%)
Engineering & Technology	51	35	86	63.2%
Arts and Humanities	18	4	22	16.2%
Social Sciences & Law	10	10	20	14.7%
Education	3	0	3	2.2%
Business & Economics	1	1	2	1.5%
Natural Sciences	1	0	1	0.7%
Prefer not to say	1	1	2	1.5%
Total	85	51	136	100.0%

Table 10: Annotator demographics.