



Sina at FigNews 2024: Multilingual Datasets Annotated with Bias and Propaganda

Lina Duaibes
Birzeit University, Palestine

Areej Jaber
Palestine Technical
University-Khadoorie

Mustafa Jarrar
Birzeit University, Palestine

Ahmad Qadi
Tamleh Center, Palestine

Mais Qandeel
Örebro University, Sweden

Contributions

❖ Multilingual Dataset

Annotated Corpus (12K FB posts) for bias and propaganda, in 5 languages.

❖ Annotation Guidelines

Annotation guidelines ensuring consistency and accuracy.

Multilingual Datasets Annotated with Bias and Propaganda

About the raw dataset

- 15,000 posts from verified Facebook accounts about the Israeli War on Gaza.
- Selected between October 6, 2023, and January 31, 2024.
- 5 languages: Arabic, Hebrew, English, French, and Hindi (3k per language).

Team

- Five experts (Authors).
- 10 Master's students specializing in Law at Birzeit University
- 7 females and 3 males.
- All native Arabic speakers with good English proficiency

Bias Guidelines

Bias: Defined as an inclination or prejudice towards or against a person or group, often unfair and based on prohibited grounds (race, religion, etc.).

Classifications:

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

Type of Bias:

- Explicit: Obvious and evident
- Implicit: Clear but not direct
- Vague: Indirect and ambiguous

Propaganda Guidelines

Propaganda: Misleading ideas/statements distorting truth or omitting facts to promote agendas.

Classifications:

- Propaganda
- Not propaganda
- Not Applicable
- Unclear

Types of propaganda

- Must be deleted: Evident harmful content
- May be deleted: Context-dependent
- Not to be deleted: Lacks harmful consequence.

Annotation Process

- Selected 200 posts for training all students in annotation
- Each student assigned 1,200 posts for annotation after training
- Dataset: 12 batches with 10,800 posts from the Main sheet, 1,200 posts from the IAA sheet

Task Participation and Results

Results

Sina team achieved:

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

Propaganda Distribution

31% of posts (3,333) annotated as "Propaganda"
66% of posts (7,084) annotated as "Not Propaganda"
The majority of propaganda is from French posts

Class	Ar	En	He	Fr	Hi	Total
Propaganda	524	679	648	809	673	3333
Not Propaganda	1484	1443	1447	1297	1413	7084
Not Applicable	48	11	17	16	24	116
Unclear	104	27	48	38	50	267
Total	2160	2160	2160	2160	2160	2160

Bias Distribution

27% of posts are biased against Palestine
63% of posts are unbiased
Most bias against Palestine originated from French posts.

Class	Ar	En	He	Fr	Hi	Total
Biased against palestine	466	514	595	807	534	2916
Biased against Israel	94	79	23	19	70	285
Biased against Both	6	7	11	6	14	44
Biased against Others	42	28	53	39	49	211
Unbiased	1371	1486	1369	1212	1386	6824
Not Applicable	49	7	17	20	25	118
Unclear	132	39	92	57	82	402
Total	2160	2160	2160	2160	2160	2160

Download



<https://sina.birzeit.edu/fada/>

Types of Bias :

Most posts with Explicit bias are in Hebrew

Types of Bias	Ar	En	He	Fr	Hi	Total
Explicit (تحيز صريح)	394	366	563	412	388	2093
Implicit (تحيز ضمني)	199	217	265	236	269	1186
Vague (تحيز مبهم)	36	37	59	52	27	211

Types of Propaganda :

Most posts classified as propaganda that must be deleted are in French (348 posts)

Class	Ar	En	He	Fr	Hi	Total
Propaganda Must be deleted	192	191	266	348	277	1274
Propaganda May be deleted	542	488	382	461	396	2059
Propaganda not to be deleted	451	422	565	648	436	2522