

# FIGNEWS 2024: Ceasefire Guidelines

## Objective and Purpose

The main goal of the shared work is to reveal and comprehend biases in news reporting about the Gaza-Israel war that occurred between 2023 and 2024. The objective of the assignment is to examine reports and associated social media postings from a viewpoint that considers several languages and cultures. This will provide a thorough examination of how other cultures and languages influence the storytelling around the war. We used NLP manual labeling and automated methods - ChatGPT API using GPT 3.5 Turbo as our Large Language Model (LLM) - to handle complex label annotations for a set of texts from news articles and social media posts in different contexts. We labeled the statements based on their presented biases, categorizing them into 7 specific groups. These labels are thoroughly described below.

## Categories: List and definitions

We present the bias categories and definitions used in our project below:

- A. Unbiased: A statement is unbiased when it presents information objectively, without favoring any particular viewpoint or agenda. This could be information with a balanced representation, factual accuracy, neutral tone and absence of personal bias.
- B. Biased against Palestine: A statement is biased against Palestine when it shows prejudice or favoritism against Palestinian interests, perspectives, or rights.
- C. Biased against Israel: A statement is biased against Palestine when it displays prejudice or favoritism against Israeli interests, perspectives, or rights.
- D. Biased against both Palestine and Israel: A statement is biased against Palestine and Israel when it demonstrates prejudice or favoritism against both Palestinian and Israeli interests, perspectives, or rights.
- E. Biased against others: A statement is biased against other when it exhibits prejudice or favoritism against parties or perspectives not directly related to Palestine or Israel but in the context of the Palestine or Israel conflict.
- F. Unclear: A statement is unclear in terms of bias when it does not clearly exhibit bias for or against any particular party or perspective in the context of the Palestine or Israel conflict.
- G. Not Applicable: A bias is not applicable to a statement that does not pertain to the Palestinian and Israeli conflict situation.

## Detailed Category Guidelines: Application criteria with examples

Below are the application criteria for each of the categories within the bias labels with examples:

- Unbiased: A statement is unbiased when it presents information objectively, without favoring any particular viewpoint or agenda. This could be information with a balanced representation, factual accuracy, neutral tone and absence of personal bias.

Text	Comment
------	---------



On Day 2 of the four-day #truce between #Israel and #Hamas, people in #Gaza are experiencing a semblance of normalcy not seen in the past few weeks. Citizens in Gaza told CGTN they hope this truce will become permanent as they are tired of the ongoing #war.	The statement appears to present a factual account of the situation in Gaza during the temporary truce between Israel and Hamas, as reported by citizens to CGTN.
"Hamas is not a state and cannot be a party to the Genocide Convention and cannot be a party to these proceedings." UK-based lawyer Vaughan Lowe, explains why South Africa does not seek a court order against Hamas at the ICJ.	The statement is a legal explanation for South Africa's decision, emphasizing the technical limitations in pursuing legal action against Hamas as a non-state entity. It does not promote a particular political agenda or viewpoint but instead offers legal reasoning. Therefore, it does not exhibit bias or propaganda.
Israel Prime Minister Benjamin Netanyahu's office said Sunday its security cabinet had approved the state of war he vowed a day earlier.	This tweet is simply telling us what happened without showing any bias in the tone.

- **Biased against Palestine:** A statement is biased against Palestine when it shows prejudice or favoritism against Palestinian interests, perspectives, or rights


Text	Comment
The head of the FBI says the bureau is "working around the clock" to stop potential terrorist attacks in the U.S. by individuals or groups inspired by Hamas' deadly assault on Israel.	The phrase "Hamas' deadly assault on Israel" shows a bias against palestine and a propaganda.
Netanyahu rejects growing calls for a cease-fire as Israel battles Hamas outside main Gaza hospital. <a href="https://arab.news/w44xp">https://arab.news/w44xp</a> := <a href="https://www.arabnews.com/node/2407206/middle-east">https://www.arabnews.com/node/2407206/middle-east</a>	the statement shows that Netanyahu is biased against palestine because it rejects ceasefire even when there is a battle going on outside Gaza hospital where people of Gaza are attacked and also phrase 'Israel battles Hamas' depicting more bias against palestine
Raf Sanchez reports live from Israel on the latest developments in Hamas' surprise military operation against Israel. "Israel, from Prime Minister Benjamin Netanyahu on downward, is clear that what is happening right now is terrorism on a very, very large scale."	Israeli Prime minister saying "what is happening right now is terrorism on a very, very large scale" as well as the tweet mentioning "Hamas' surprise military operation against Israel" shows a bias towards Israel and is a propaganda.

- **Biased against Israel:** A statement is biased against Palestine when it displays prejudice or favoritism against Israeli interests, perspectives, or rights.

Text	Comment
#Urgent   Head of the Hamas Political Bureau: - With greater determination to continue the path of the martyrs, we mourn the Mujahid leader	Shows bias against Israel by referring to the occupation's assassination as a complete terrorist act and a violation of Lebanon's sovereignty, while

#Saleh_Al-Arouri. - We mourn Commander Saleh Al-Arouri and the Qassam leaders Samir Fandi and Azzam Al-Aqra'. The occupation's assassination of Al-Arouri and his fellow movement leaders is a complete terrorist act and a violation of #Lebanon's sovereignty. The occupation will not succeed in breaking the will of steadfastness and resistance among our people and their valiant resistance. #Gaza_War	praising the resistance and steadfastness of the people.
Is Israel really rounding up ' Hamas fighters' or is it just humiliating Palestinian civilian men and boys and broadcasting it to the world to dehumanise them? – #AJOpinion by Marc Owen Jones   : <a href="https://www.aljazeera.com/opinions/2023/12/10/fact-or-fiction-is-israel-really-rounding-up-hamas">aje.io/qkwepo=:https://www.aljazeera.com/opinions/2023/12/10/fact-or-fiction-is-israel-really-rounding-up-hamas</a>	This statement suggests that Isreal is "humiliating Palestinian civilian men and boys and broadcasting it to the world to dehumanise them," showing a bias. It also shows how Isreal is spreading propaganda by "broadcasting." This sentence, itself, is making us question Isreal's motives and creates a propaganda.
Benjamin Netanyahu refuses to consider a ceasefire. The Israeli army speaks of fierce fighting in the Gaza Strip, Hamas has counted more than 8,500 dead since October 7, it is also an information war that we will analyze with our guests	The statement portrays Benjamin Netanyahu in a negative light by mentioning his refusal to consider a ceasefire. It also highlights the high death toll on the Palestinian side without providing context or mentioning any Israeli casualties, showing bias against Israel and in favor of Palestine.

- Biased against both Palestine and Israel: A statement is biased against Palestine and Isreal when it demonstrates prejudice or favoritism against both Palestinian and Israeli interests, perspectives, or rights

Text	Comment
College campuses across the United States have been at a fever pitch in the past week over statements from students, professors and administrators related to Hamas' attack on Israel and the ensuing war in Gaza.	Talks about both, bias against palestine based on phrase 'the ensuing war in Gaza' and bias against israel 'Hamas' attack on Israel', not propaganda because it just mentions the situation in us right now
 In the Law Committee, during our debates on credits dedicated to immigration, I wanted to point out some obvious facts to NUPES and the Macronists: 👉 The National Rally has no lesson in humanism to receive from the from friends of Hamas. 👉 He has no lessons in competence to receive from a majority whose migration policy is rejected by 80% of French people.	The statement criticizes both the National Rally and the majority, highlighting their shortcomings and implying bias against both parties.
Macron fears neither ridicule nor inconsistency. For fear of riots, he serves two peoples. One day he supports Israel's right to defend itself; the next day, he no longer knows why Israel had the right to defend itself, orders everything to stop and offers millions to Gaza.	The statement criticizes Macron for inconsistency in his stance towards Israel and Gaza, suggesting bias against both Palestine and Israel.

- **Biased against others:** A statement is biased against other when it exhibits prejudice or favoritism against parties or perspectives not directly related to Palestine or Israel but in the context of the Palestine or Israel conflict.

Text	Comment
America vs Iran in Israel! Gaza destroyed! Now ground attack! #FullandFinal   #WarCoverageOnTV9   #Israel HamasConflict	The statement mentions America and Iran in Israel, implying involvement of other countries in the conflict. It also highlights Gaza being destroyed and a ground attack, showing a biased against others.
For many, UN Women's belated condemnation of Hamas' use of sexual violence on October 7 fell short, writes Ben Samuels	The statement criticizes UN Women for not condemning Hamas' use of sexual violence, showing bias against Hamas.
Russian missile system to Hezbollah, big news amid Israel war #IsraelHamasWar #GAZA #IsraelPalestineWar	Mentions the transfer of a Russian missile system to Hezbollah during the Israel war, indicating bias against others by criticizing entities other than Israel or Palestine

- **Unclear:** A statement is unclear in terms of bias when it does not clearly exhibit bias for or against any particular party or perspective in the context of the Palestine or Israel conflict.

Text	Comment
'They treated us very nicely.'	The statement lacks context and does not specify who 'they' refers to, making it unclear.
LIVE   Who is Laden with wheel chair? How will Gaza Strip become 'soil'? See 'Public Interest'	Limited context and unclear meaning, does not show bias towards any party
What's going to happen next in the Israel-Hamas war? #SpecialReport #ATVideo #IsraelGazaWar   Anjana Om Kashyap	Limited context, just a question about the future of the conflict without showing bias towards any party.

- **Not Applicable:** A bias is not applicable to a statement that does not pertain to the Palestinian and Israeli conflict situation.

Text	Comment
Al Kass Channel Council A tour of Souq Waqif observes great public enthusiasm. #Come_Asia   #AsiaCup2023 #Qatar2023   #AsianCup2023 #Asian_Cup_Cup_Channels #See_the_Cup	No mention of bias towards Palestine, Israel, or any other party in the conflict. The statement is focused on observing public enthusiasm at Souq Waqif.
"When freedom of speech becomes freedom of hate it becomes a whole different thing."	no context related to any bias against Palestine or Israel or other groups , general vague statement
Crazy: The freed kidnapper managed to escape from captivity - and was caught again	The statement is about a kidnapper escaping and being caught again, with no mention of bias

	towards any party in the Israel-Palestine conflict.
--	---

## Common mistakes

We now present a list of common labeling mistakes with reasons for the errors.

Text	Correct label	Possible mistake	Comment
A Controversial Video Goes Viral on Social Media: Women Laugh and Take Selfies in the Background of a Kidnapping, Displaying Disrespectful Gestures. #Israel #HamasWar #IsraelUnderAttack #Gaza #Palestinians #IsraelPalestineWar #Gaza #IsraelFightsTerror #IndiaStandWithIsrael #telaviv #Hezbollah #FPJ	Biased against Palestine	Biased	Since these texts are from social media, which rely on hashtags, it is important to analyze them too. In this example, if without the hashtag, this text appears to be unclear however, after reading the hashtag “#IsraelUnderAttack” and “#IsraelFightsTerror,” we can tell that this post hints a bias toward Israel.
US President Joe Biden says both Hamas and Russia’s Vladimir Putin “want to completely annihilate a neighboring democracy.” <a href="https://trib.al/UAfb8eA=:https://www.bloomberg.com/news/articles/2023-10-20/biden-casts-russia-hamas-as-parallel-threats-to-democracy">https://trib.al/UAfb8eA=:https://www.bloomberg.com/news/articles/2023-10-20/biden-casts-russia-hamas-as-parallel-threats-to-democracy</a>	Biased against Palestine	Biased against others	A potential mistake from someone who doesn’t understand the context of the events, might not know “Hamas” here, on broader terms refers to Palestine but rather think of them as a separate party and label this text as biased against others.
Hummus in tahini, which became one of the symbols of the national-culinary conflict, has a common heritage in the Middle East. To attribute it only to Palestinians or to Israelis is almost a contradiction in terms   Opinion	Not Applicable	Unbiased	A misunderstanding of the term and context in the text. A person who is not much aware of the situation could lead to incorrect labeling
Watch IDF Spokesperson RAdm. Daniel Hagari walk through one of Hamas’ subterranean terrorist tunnels — only to exit in Gaza’s Rantisi hospital on the other side. Inside these tunnels, Hamas terrorists hide, operate and hold	Biased against Palestine	Unbiased	Mentions that Hamas is a terrorist and that it holds Israeli hostage as part of this war.

Israeli hostages against their will Israel Defense Forces דובר צה"ל IDF Spokesperson			
Hamas delays release of Israeli hostages on second day of its truce with Israel in dispute over aid and Palestinian prisoners. Follow live updates.	Biased against both Palestine and Israel	Biased against Palestine	Hamas delayed the release of Israeli hostages and Israel is in disputer over aid and prisoners

## Process Outline: Step-by-step annotation process and tools

### Manual annotations

- 1) The team first established guidelines for each label.
- 2) Then, together completed 40 annotations with well-structured comments from the main sheet following the guidelines. This helped us identify any errors or additions to the guidelines together. We finalized the guidelines.
- 3) The members then used the guidelines to complete the inter-annotator agreement (IAA) and checked our results to make sure they had high rate of similarities.
- 4) Next, we completed 20 annotations each from the main sheet, following our finalized guidelines.
- 5) Overall, we had 100 manual annotations in total, which we would use to train and test our model.

### Annotation Process Using ChatGPT with Prompt Engineering

Our methodology involves using Retrieval-Augmented Generation (RAG) and a large language model (LLM) to automate the labeling process. Initially, we manually labeled 100 statements and provided detailed comments to serve as in-context learning examples for the model. We used the GPT-3.5 Turbo API from OpenAI, which required crafting a well-structured prompt that included the 100 examples to guide the model in accurately labeling statements. The prompt also included explicit definitions for terms like Hamas, IDF, and Netanyahu to help the model understand the context and correctly identify bias. For example:

**Hamas:** The Islamic Resistance Movement, known as Hamas, represents Palestine and its people. Any mention of Hamas typically relates to Palestine.

**Israel-Hamas:** Any statement that mentions Israel-Hamas basically is a balanced view and is just referring to the war which means it doesn't indicate any bias on its own and is just talking about the war in general terms from both sides.

**Our simplified version of the Prompt:** *"You are an expert in labeling. You will be given a set of statement in English which you would have to label based on two guidelines (Bias and propaganda):*

*Corpus comprising news article headlines and advertising posts about them from Facebook (henceforth, posts)*

*Bias Guidelines:*

*Unbiased: A statement is unbiased when it presents information objectively, without favoring any particular viewpoint or agenda. This could be information with a balanced representation, factual accuracy, neutral tone and absence of personal bias.*

*Biased against Palestine: A statement is biased against Palestine when it shows prejudice or favoritism against Palestinian interests, perspectives, or rights.*

*Biased against Israel: A statement is biased against Palestine when it displays prejudice or favoritism against Israeli interests, perspectives, or rights.*

*Biased against both Palestine and Israel: A statement is biased against Palestine and Isreal when it demonstrates prejudice or favoritism against both Palestinian and Israeli interests, perspectives, or rights.*

*Biased against others: A statement is biased against other when it exhibits prejudice or favoritism against parties or perspectives not directly related to Palestine or Israel but in the context of the Palestine or Israel conflict.*

*Unclear: A statement is unclear in terms of bias when it does not clearly exhibit bias for or against any particular party or perspective in the context of the Palestine or Israel conflict.*

*Not Applicable: A bias is not applicable to a statement that does not pertain to the Palestinian and Israeli conflict situation.*

*You should also include a short comment explaining your choice of bias type and propaganda category. Each comment should include maximum 30 words. The output should be organized into three columns in an excel sheet: the first column should include the statement, the second column should state the specific bias type from the guidelines, the third column for the specific propaganda type, and the fourth column for your comments.*

*For your context, below are some definitions mentioned within the set of tweets:*

*Hamas: Hamas stands for the Islamic Resistance Movement and in Arabic means “zeal”.*

*The group politically controls the Gaza Strip, a territory of about 365sq km (141sq miles) that is home to more than two million people but is blockaded by Israel.*

*Hamas has been in power in the Gaza Strip since 2007 after a brief war against Fatah forces loyal to President Mahmoud Abbas, the head of the Palestinian Authority and Palestine Liberation Organization (PLO).*

*Benjamin "Bibi" Netanyahu is an Israeli politician who has been serving as the prime minister of Israel since 2022, having previously held the office from 1996 to 1999 and again from 2009 to 2021. He is the chairman of the Likud party"*

We set the model's temperature to 0.4, which adjusts the randomness of the model's responses. With this value, the model produced consistent results. The prompt defined bias types and terms that the model needed to understand and included in-context learning examples.

The model processed data and gave the output in about 2.5 hours approximately and we instructed the model to provide both labels and comments. The comments were crucial to understand the model's reasoning behind each label. Using GPT-3.5 Turbo cost us about \$5 to run the entire model, and the output was delivered in an Excel file.

The output was mostly accurate, but there were some errors. The model incorrectly labeled a statement as "bias against Hamas" once and misclassified seven statements as "bias against Israel, bias against others," or "bias against Palestine, bias against others." These errors occurred because the sentences had a bias against multiple groups, including Palestine/Israel. We manually reviewed these statements and corrected the labels to reflect the majority bias against a specific group. We also used 40% of our manually annotated data for testing the outputs.

After these manual corrections, a random check of the output showed that the model achieved 98% accuracy, making this process effective for our labeling needs.

## **Quality Standards**

All the manually labeled data was checked by all the annotators to make sure it was accurate and consistent as this is what would train and test our model. Any errors were followed up with team discussions to make sure we all mutually agreed on all the labels.

The outputs presented by the model were also reviewed by us through a random check to make sure it retrieved us with accurate results. We kept running the model reviewing the results, and making necessary changes (mostly providing more context to the LLM) until we were satisfied with the outputs. The model was also tested with 40% of our manually labeled data to test for accuracy.

## **Handle Ambiguities**

The initial testing stage of our model showed that the model presented ambiguities or error due to a lack of context understanding. To avoid these, we provided our prompt with details and context (see the prompt above). Also, since our model pays close attention to spelling, many



ambiguities from the texts were easily avoided (eg. Hamas and “Hummus” as presented in the common mistakes table).

### **Ensure Consistency**

Our team manually checked the consistencies among all our manual annotations and held review sessions very frequently to make sure we agreed with all the labels.

### **Ethical Considerations**

Our model utilizes ChatGPT, which initially shows a bias towards Israel. However, we set a blank slate by asking it to label our texts based on the provided context only. We also set the model's temperature to 0.4 to avoid hallucination.

### **Training and Support**

Since all the annotators have been working on this project from the start, we trained and established the guidelines together. All the review sessions were done together. In terms of training the model, we provided it with training data we all agreed on and manually completed.

### **Review, Feedback and, Update**

We established the guidelines together and annotated some of the statements together to make sure we provided feedback, and reviewed and updated the guidelines together. Furthermore, constantly held meetings to make sure we were frequently meeting to check the annotations, provide feedback to each other, and update the guidelines accordingly. At the end, we randomly checked all the annotations together and found 99% accuracy and consistency.

### **Revision Work**

- Shorten the introduction and add a few sentences to include the title info.
- Inclusion of specific examples of the guidelines in Guidelines creation and data annotation process section.
- Detailed examples of guidelines in the main text.
- Include comparative analysis with related work for the results
- Format doesn't follow ACL guidelines.
- Table descriptions show the following ACL guidelines.

**Unbiased**: A statement is unbiased when it presents information objectively, without favoring any particular viewpoint or agenda. This could be information with a balanced representation, factual accuracy, neutral tone and absence of personal bias.

Israel Prime Minister Benjamin Netanyahu's office said Sunday its security cabinet had approved the state of war he vowed a day earlier.

\textbf{Biased against Palestine}: A statement is biased against Palestine when it shows prejudice or favoritism against Palestinian interests, perspectives, or rights.

Israel is expected to launch a ground invasion into Gaza after Hamas terror attacks.

**Adherence to Title:** We agree with your observation and we have clarified in the introduction how the paper adheres to the title, ensuring the scope and objectives are clearly aligned.

**Annotation Methodology and Examples:** Yes, we have included detailed descriptions of the guidelines creation and data annotation process. We also provided specific examples to enhance clarity.

**Guidelines Examples::** Yes, we have provided a few detailed examples in the main text. The remaining examples are included in the appendix section for completeness.

**Task Participation and Results:** We appreciate your feedback and have included more comparative analysis with related work to provide valuable insights and contextualize our findings.

**Comparative Analysis:** Yes, we have included comparative analysis with previous related work to provide valuable insights and contextualize our findings.

**Additional Format and Requirement Issues: Format Compliance::** Yes, we have revised the format to comply with ACL guidelines.

**Line Numbering:** Yes, we have added line numbers to comply with the submission guidelines.

**Tables Description:** Yes, we have ensured that all table descriptions now follow the ACL format.

**Appendices Titles:** We understand the importance of clarity and completeness. However, the appendix section contains the entire prompt used in the LLM model, serving as a single, comprehensive illustration. Adding subheadings within this prompt might disrupt the clarity and coherence of how it is presented to our model.

**Corpus Information Placement:** Yes, we have appropriately placed the corpus information, moving it to an endnote as suggested.