Integrating Argumentation Features for Enhanced Propaganda Detection in Arabic Narratives on the Israeli War on Gaza

Sara Nabhani[1,2], Claudia Borg[1], Kurt Micallef[1], Khalid Al-Khatib[2]

[1]Department of Artificial Intelligence, University of Malta
{sara.nabhani.23, claudia.borg, kurt.micallef}@um.edu.mt
[2]Department of Computational Linguistics and Society, University of Groningen
{s.nabhani, khalid.alkhatib}@rug.nl

Abstract

Propaganda significantly shapes public opinion, especially in conflict-driven contexts like the Israeli-Palestinian conflict. This study explores the integration of argumentation features, such as claims, premises, and major claims, into machine learning models to enhance the detection of propaganda techniques in Arabic media. By leveraging datasets annotated with fine-grained propaganda techniques and employing cross-lingual and multilingual NLP methods, along with GPT-4-based annotations, we demonstrate consistent performance improvements. A qualitative analysis of Arabic media narratives on the Israeli war on Gaza further reveals the model's capability to identify diverse rhetorical strategies, offering insights into the dynamics of propaganda. These findings emphasize the potential of combining NLP with argumentation features to foster transparency and informed discourse in politically charged settings.

1 Introduction

Propaganda is a form of communication aimed at influencing attitudes and behaviors by presenting one-sided or misleading information. It often relies on emotional appeals rather than rational argumentation to manipulate public perception and advance specific agendas or ideologies.

In the digital era, the rise of social media has amplified the spread of propaganda, enabling its rapid dissemination to global audiences with little oversight. This has heightened its potential impact, as seen in events like the 2016 U.S. Presidential Election (Ali and ul abdin, 2021) and during the COVID-19 pandemic (Broniatowski et al., 2020), where social media platforms were used to polarize opinions and undermine trust in democratic institutions.

The detection of propaganda is especially critical in conflict-driven contexts, such as the narratives surrounding the Israeli war on Gaza. These narratives often employ polarizing rhetoric, emotionally charged language, and manipulative techniques to shape public opinion and justify political or military actions. Arabic media, both traditional and digital, plays a central role in constructing these narratives, given the geopolitical significance of the Arabic-speaking world. In such contexts, propaganda can be a powerful tool for inciting violence, manipulating perceptions, and influencing international discourse. However, detecting propaganda in Arabic poses unique challenges due to the language's rich morphology, diverse dialects, and limited annotated datasets.

Natural Language Processing (NLP) offers a promising avenue for automating propaganda detection by analyzing linguistic patterns and rhetorical cues . While significant progress has been made in high-resource languages like English, relatively little research has focused on Arabic. This disparity highlights the need for approaches tailored to Arabic's linguistic and cultural characteristics.

A promising direction for enhancing propaganda detection is the integration of argumentation features, such as claims and premises. Propaganda and argumentation often share a structural foundation: both involve presenting claims supported by reasoning. However, propaganda diverges by infusing these structures with emotionally charged content designed to manipulate public sentiment (Nettel and Roque, 2012). By identifying argumentation components within texts, it becomes possible to analyze how propaganda leverages these structures to influence audiences, distinguishing between logical persuasion and manipulative communication.

In this work, we aim to improve Arabic propaganda detection by integrating argumentation features into NLP models. We then apply the enhanced models to analyze narratives from Arabic media covering the Israeli war on Gaza. The code used in this study is available at our GitHub repository.[1]

[1]https://github.com/saranabhani/prop-arg

Proceedings of the Proceedings of the 1st International Workshop on Nakba Narratives as Language Resources, pages 127–149
January 20, 2025.

2 Related Work

2.1 Propaganda Detection in Arabic Texts

A shared task on Propaganda Detection in Arabic was organized at the WANLP 2022 workshop (Alam et al., 2022) to address the notable absence of research on Arabic language propaganda detection. In this shared task, one submission (Hussein et al., 2022) applied basic preprocessing steps like normalization and transformed the data into the BIO format to represent data spans within tweets accurately. They adopted a transfer learning approach by employing the Marefa-NER model, a pre-trained template designed for Named Entity Recognition (NER), demonstrating the model's adaptability to this propaganda detection task.

Building on the momentum of the WANLP 2022 shared task on Propaganda Detection in Arabic (Alam et al., 2022), the organizers introduced the ArAIEval shared task[2] (Hasanain et al., 2023) focusing on two critical areas: persuasion technique and disinformation detection in tweets and news articles. The top submission (Lamsiyah et al., 2023) achieved first place with a streamlined approach centered around a BERT-based Arabic pre-trained language model encoder coupled with a singular, efficiently structured classifier. In their exploration of input text encoding, the team assessed the performance of three BERT-based Arabic pre-trained language models: ARBERTv2 (Abdul-Mageed et al., 2021), MARBERTv2 (Abdul-Mageed et al., 2021), and AraBERT-large (Antoun et al., 2020). The AraBERT encoder was selected, and the model was trained using an asymmetric multi-label loss.

Capitalizing on the progress achieved by the ArAIEval shared task, the 2024 edition (Hasanain et al., 2024) continued to advance the field of propaganda detection in Arabic text. Task 1 of the shared task focused on Unimodal Propaganda Detection, specifically targeting the identification of persuasive techniques within tweets and news articles written in Arabic. The dataset used for this task comprised tweets derived from Arabic news sources on Twitter, along with news paragraphs sourced from the AraFacts dataset (Sheikh Ali et al., 2021). The annotation process for this dataset involved labeling text snippets with a set of 23 persuasion techniques, building on the work of Piskorski et al. (2023). The top submission for this task came from Labib et al. (2024), which achieved the highest F1 score by integrating data augmentation techniques with model fine-tuning. Their approach involved leveraging a pre-trained Arabic-BERT model (Safaya et al., 2020), which was specifically fine-tuned on the task's annotated data. To address the challenge of class imbalance, the team employed data augmentation strategies such as synonym replacement, which enhanced the model's ability to generalize across different types of persuasive techniques. Another strong submission was from Riyadh and Nabhani (2024), who took advantage of a multilingual BERT model (mBERT) (Devlin et al., 2019) to capture the complexities of Arabic text. Their approach was distinctive in its focus on experimenting with different hidden layers of the model to determine the most effective layer for the task.

Overall, the recent advancements in propaganda detection in Arabic text have predominantly relied on fine-tuning transformer-based architectures and leveraging data augmentation techniques.

2.2 Contextual Features in Propaganda Detection

Relatively few studies have explored the integration of contextual features to enhance the performance of propaganda detection systems. A notable exceptions involve the addition of discourse features to token embeddings, which has shown potential for improving the accuracy of propaganda detection. This study by Chernyavskiy et al. (2024) explored the integration of discourse features into token embeddings to enhance the detection of propaganda in English and Russian. For this they used the dataset from SemEval-2023 (Piskorski et al., 2023). Their approach involved conducting a discourse analysis of the text to identify higher-level organizational structures utilizing the Two-Stage discourse parser (Wang et al., 2017). By embedding these discourse features directly into the token representations, the model gained a richer understanding of the text's structure, which proved beneficial in identifying propagandistic content.

This study highlights the significant impact of incorporating contextual features into token embeddings. This approach provides models with a deeper understanding of the context surrounding propaganda, beyond just the surface-level content of the text. While research in this area is still relatively sparse, the positive outcomes from these studies support the potential of our methodology in this work, suggesting that further exploration could lead to significant advancements in propaganda de-

[2] https://araieval.gitlab.io/

tection.

3 Data

3.1 Propaganda Detection Dataset

For the propaganda detection task, we utilized the dataset provided by the ArAIEval 2024 shared task on propaganda detection in Arabic text (Hasanain et al., 2024), specifically focusing on Task 1: Unimodal (Text) Propagandistic Technique Detection. This dataset encompasses two primary text genres: tweets and paragraphs extracted from Arabic news articles. Details regarding the data collection and annotation processes are thoroughly documented in the shared task paper (Hasanain et al., 2024). The dataset is publicly accessible via the ArAIEval GitLab repository.[3]

The dataset is pre-split into training, validation, and test sets, which were directly utilized in this work without modification. Each entry in the dataset contains a unique identifier, the raw text (either a tweet or a news paragraph), and annotations describing the propaganda techniques identified within specific spans of text. Each annotation includes the technique name, the exact text span where the technique occurs, and the character positions marking the start and end of the span. Text spans can be associated with multiple propaganda techniques, and overlapping spans are common.

The dataset includes 23 fine-grained propaganda techniques, derived from the taxonomy proposed by Piskorski et al. (2023). Detailed explanations of each technique, as defined in Piskorski et al. (2023), can be found in Appendix A.

The dataset's structure allows for a comprehensive analysis of propaganda in Arabic texts, accommodating both sequence labeling (to identify specific spans of propaganda) and multilabel classification (to categorize the techniques used).

Table 1 presents detailed statistics, including the sizes of the training, validation, and test sets and the total number of tokens. Figures 1a and 1b in Appendix B visualize the distribution of propaganda techniques across the datasets. Techniques such as Loaded Language and Name Calling are the most frequent, while others, like Straw Man and Guilt by Association, appear less commonly. The label distribution across the training, validation, and test sets is relatively consistent, despite the uneven number of different labels.

[3]https://gitlab.com/araieval/araieval_arabicnlp24

	Train	**Dev**	**Test**
# Documents	6,997	921	1,046
# Tokens	228,373	27,867	35,204
Avg. Tokens/Doc	32.63	30.25	33.65
Unique Tokens	59,193	13,443	16,108

Table 1: Propaganda dataset statistics

3.2 Argumentation Mining Dataset

To incorporate argumentation features into our study, we utilized the Persuasive Essays (PE) corpus by Stab and Gurevych (2017), as no suitable Arabic datasets aligned with our requirements. This English-language resource, widely used in cross-lingual argumentation tasks, comprises 402 essays randomly selected from essayforum.com, each accompanied by writing prompts and annotated with key argumentation components. These components include: Major Claim, representing the central argument typically introduced in the introduction and reinforced in the conclusion; Claim, which supports or challenges the major claim by addressing specific aspects or perspectives; and Premise, consisting of evidence or reasons that justify a claim and explain its validity. The corpus is pre-divided into training and test sets, which we used without modification. Table 2 provides detailed statistics about the corpus. By adapting this robust English dataset through a cross-lingual framework, we aim to extend its applicability to Arabic, leveraging its detailed annotations to enhance our study.

	Train	**Test**	**Total**
# Essays	322	80	402
# Paragraphs	1,786	449	2,235
# Tokens	118,645	29,537	148,182
MajorClaim	598	153	751
Claim	1,202	304	1,506
Premise	3,023	809	3,832

Table 2: Argumentation dataset statistics

3.3 Analysis Dataset: News Media Narratives on the Israeli War on Gaza

The dataset used for analyzing news media narratives about the Israeli war on Gaza originates from the FIGNEWS shared task (Zaghouani et al., 2024). This initiative focused on the early stages of the Israel-Gaza conflict, curating a multilingual corpus covering five languages: Arabic, English, French, Hebrew, and Hindi.

The dataset was annotated with multiple layers, including bias labels ("biased against Palestine", "biased against Israel", "unbiased") and propaganda labels ("propaganda", "not propaganda").

The qualitative analysis for this work (Section 8) utilized a sample of 17 examples with Arabic source language from the top-performing system in the shared task, developed by team NLPColab (Abdul Rauf et al., 2024).

4 Baseline for Propaganda Detection

In this study, we use the best-performing system from the ArAIEval 2024 shared task on propaganda detection in Arabic texts by Labib et al. (2024) as our baseline. This system, built on Arabic-BERT (Safaya et al., 2020), achieved an F1 score of 0.2995 by fine-tuning for detecting propagandistic spans and classifying them into 23 techniques. Key features include the BIO tagging scheme for accurate span identification and data augmentation to address class imbalance for less frequent techniques.

While this system was not originally developed as a baseline, we adopt it in this role for our study. Its performance in the shared task makes it an ideal reference point for evaluating the improvements introduced by our approach.

5 Proposed Methodology

This study investigates the enhancement of propaganda detection models by integrating argumentation features. Argumentation features, such as *Major Claim*, *Claim*, and *Premise*, provide a structured representation of the persuasive elements within a text. By leveraging the overlap between argumentation and propaganda, we aim to enrich the model's understanding of the underlying rhetorical strategies.

5.1 Model Architecture

The proposed model builds on a transformer-based architecture with AraBERTv2 (Antoun et al., 2020) as the backbone. This pre-trained model generates rich contextual embeddings for each token, capturing linguistic characteristics in Arabic text. To incorporate argumentation features, we augment these embeddings with additional input, as described below.

Input Representation Each token in the input text is represented by a combination of contextual embeddings and argumentation features. The token embeddings, derived from AraBERTv2, capture the linguistic and contextual information of each token. Additionally, a one-hot encoded vector of length four represents argumentation features, assigning each token one of four values: *Major Claim*, *Claim*, *Premise*, or *None*. These argumentation features, generated by an argumentation analyzer, are concatenated with the token embeddings to create a richer and more comprehensive representation.

Output Representation The model is designed for multi-label classification at the token level, where each token is assigned a binary vector representing the propaganda labels. The vector length corresponds to the number of propaganda techniques considered in the study (23 techniques). A value of 1 in the vector indicates the presence of a specific propaganda technique, while a value of 0 denotes its absence.

5.2 Model Workflow

The proposed model's workflow begins with embedding generation, where the input text is tokenized and processed through AraBERTv2 to produce contextual embeddings. These embeddings are then augmented with argumentation features, which are concatenated to enrich the representation of each token. The enhanced embeddings are passed through a classification layer to compute probabilities for various propaganda techniques. Finally, propagandistic spans are identified by grouping consecutive tokens with identical labels.

6 Argumentation Annotation Methodology

To generate argumentation annotations, we employed two primary approaches:

1. **GPT-4 Prompting**: This method involved using GPT-4 to automatically annotate the data.

2. **Trained Argumentation Model**: A dedicated argumentation model was developed and trained on the Persuasive Essays (PE) argumentation data. Once trained, this model was applied to annotate the propaganda dataset.

By implementing these two approaches, we aimed to compare their effectiveness in augmenting the propaganda detection task with argumentation features. This comparison allowed us to evaluate and determine the optimal method for integrating argumentation annotations into the overall framework.

6.1 Argumentation Annotation with GPT-4

We utilized GPT-4o,[4] an advanced variant of GPT-4, to annotate the propaganda dataset with argumentation features. This approach leverages GPT-4o's ability to adapt without extensive task-specific training, serving as both an evaluation of its effectiveness and a baseline for comparison with trained argumentation models.

Prompt Design and Testing We experimented with different prompting strategies to guide GPT-4o in classifying spans as *Major Claim*, *Claim*, *Premise*, or *None*, using a sample of 10 sentences from the training data. Both sentence-level and word-level approaches were tested, with sentence-level prompts generally producing cleaner and more accurate annotations. In contrast, word-level prompts faced challenges such as fragmented spans and inconsistent labeling, requiring significant post-processing. Additionally, an Arabic, human-translated, version of the most effective sentence-level prompt was tested, maintaining clarity but necessitating further validation through extensive post-processing.

6.2 Argumentation Model Development

To train an argumentation analysis model for Arabic texts, we explored two strategies: monolingual modeling and multilingual modeling. These strategies effectively leveraged annotated English resources while addressing the scarcity of Arabic argumentation datasets.

Monolingual Modeling Monolingual modeling involved using English argumentation data and applying translation techniques to bridge the gap between English and Arabic. Two approaches were employed:

Translate-Train The Translate-Train approach involved translating the English Persuasive Essays (PE) argumentation dataset into Arabic. Annotation projection techniques were then applied to transfer English annotations onto the translated Arabic text, ensuring the preservation of argumentative structures. The resulting Arabic dataset was used to fine-tune a model based on AraBERTv2 (Antoun et al., 2020).

Translate-Test In this approach, RoBERTa-large (Liu et al., 2019), trained on the English PE dataset, was utilized for argumentation detection. Arabic propaganda texts were translated into English, allowing the English-trained model to annotate the translated texts. The resulting annotations were projected back onto the original Arabic texts using alignment techniques. This approach avoided direct training on Arabic data while still enabling argumentation detection.

Multilingual Modeling Multilingual modeling leveraged pre-trained multilingual transformer models, XLM-RoBERTa-large (Conneau et al., 2019), to perform argumentation detection across languages without requiring extensive annotated resources in Arabic.

Zero-Shot Multilingual Modeling The zero-shot approach involved training a multilingual model on the English PE dataset and directly applying it to Arabic propaganda texts.

Translate-Train Multilingual Modeling The Translate-Train Multilingual approach extended the Translate-Train method by combining English PE data and its translated Arabic counterpart into a single training dataset. This approach exposed the multilingual model to both languages, allowing it to learn language-specific features alongside shared linguistic patterns.

Translation and Annotation Projection For both Translate-Train and Translate-Test approaches, translation and annotation projection were critical components. ***Translation Methods:*** Two machine translation tools were employed: (1) NLLB 1.3B (Team et al., 2022), a multilingual translation model designed to handle diverse languages, including low-resource ones, and (2) Google Translate, which allowed for comparison of translation quality's impact on model performance. ***Annotation Projection:*** FastAlign (Dyer et al., 2013), a statistical word alignment tool, was used to align annotations between English and Arabic, preserving argumentative structures across translations.

By combining translation methods, annotation projection, and diverse models, our framework effectively addressed the challenges of generating argumentation annotations for Arabic texts, enabling argumentation detection in resource-constrained settings.

Mitigating the Impact of Annotation and Translation Errors The Translate-Train and Translate-Test models rely heavily on automatic translation

[4] Accessed in July 2024

and annotation projection, both of which can introduce errors that affect model performance. To address these challenges, we conducted targeted investigations to evaluate and mitigate the impact of these errors.

Annotation Projection Errors To understand the effect of annotation projection errors, we manually corrected samples of 100 and 200 instances from the training data. These corrected annotations were used to assess their impact on the performance of both the argumentation detection and the propaganda detection models. Due to the labor-intensive nature of manual corrections, this effort was focused on the Translate-Train Monolingual approach.

Translation Quality Errors Translation quality was identified as a critical factor influencing model effectiveness, particularly in the Translate-Test approach. Inspired by the findings of Artetxe et al. (2023), two key strategies were implemented to mitigate the impact of translation errors. First, **Domain Adaptation** was applied by fine-tuning the machine translation model on domain-specific data, ensuring translations better aligned with the characteristics of the argumentation detection task. Second, **Training Data Adaptation** involved augmenting the training data by translating it into Arabic and then back-translating it into English, incorporating the back-translated content to expose the model to the variability introduced by translation. These strategies highlighted the sensitivity of the Translate-Test approach to translation quality.

7 Propaganda Detection Evaluation and Discussion

The effectiveness of incorporating argumentation features into the propaganda detection task was evaluated using various approaches, including cross-lingual, multilingual, and GPT-4-based annotation methods. Table 3 summarizes the Micro F1 scores for the development and test sets, highlighting the impact of these methods on performance compared to the baseline.

7.1 Cross-Lingual Approaches

Translate-Test Using Google Translate, this method achieved a Micro F1 score of 0.3948 on the development set and 0.3978 on the test set. The NLLB translation model performed comparably, with scores of 0.3981 and 0.4024 on the development and test sets, respectively. Training data adaptation improved performance for Google Translate, reaching 0.4089 on the development set and 0.4018 on the test set. However, domain adaptation reduced performance, highlighting that this approach was not beneficial in mitigating poor translation quality.

Translate-Train Monolingual Using the NLLB-translated dataset improved performance to 0.3695 on the development set and 0.3701 on the test set. Manual corrections of annotation projection for 100 and 200 samples boosted scores on the test set to 0.3952 and 0.3947, respectively, underscoring the importance of high-quality annotation alignment.

7.2 Multilingual Approaches

Zero-Shot Multilingual achieved a Micro F1 score of 0.3981 and 0.3930 on the development and test sets, respectively. This result indicates that the model could generalize across languages, although linguistic differences between English and Arabic pose challenges.

Translate-Train Multilingual using Google Translate, achieved Micro F1 scores of 0.4033 and 0.3931 on the development and test sets, respectively. NLLB yielded similar results, with scores of 0.3988 and 0.3929. These results demonstrate a very marginal improvement over the Zero-Shot Multilingual model, indicating the benefit of multilingual exposure during training is very limited.

7.3 GPT-4 Annotation Approach

The GPT-4-based approach, using an English prompt to annotate the propaganda dataset with argumentation features, achieved the highest Micro F1 scores of 0.4077 on the development set and 0.4025 on the test set. This method demonstrated consistent performance across both datasets, outperforming other approaches.

7.4 Discussion

The results reveal several key findings. All methods incorporating argumentation features outperformed the baseline Micro F1 score of 0.2995, demonstrating the effectiveness of integrating argumentation information into propaganda detection models. Translation quality played a crucial role, as the Translate-Test approaches showed better performance with higher-quality translations, although gains were limited without adaptation techniques. The accuracy of annotation projection was also

Approach	MT Model	Adaptation	#Corrected Annotation	Micro F1 Dev	Micro F1 Test
Baseline	-	-	-	-	**0.2995**
Zero-Shot Multilingual	-	-	-	**0.3981**	**0.3930**
Translate-Test	Google Translate	-	0	0.3948	0.3978
	Google Translate	Training Data	0	**0.4089**	0.4018
	NLLB	-	0	0.3981	**0.4024**
	NLLB	Training Data	0	0.3918	0.4006
	NLLB	Domain	0	0.3773	0.3799
Translate-Train Monolingual	NLLB	-	0	0.3695	0.3701
	NLLB	-	100	0.3889	**0.3952**
	NLLB	-	200	**0.4033**	0.3947
Translate-Train Multilingual	Google Translate	-	0	**0.4033**	**0.3931**
	NLLB	-	0	0.3988	0.3929
GPT-4 - Prompt6(AR)	-	-	-	0.4004	0.3914
GPT-4 - Prompt1(EN)	-	-	-	**0.4077**	**0.4025**

Table 3: F1 Scores of Propaganda Detection Models with Argumentation Feature Augmentation Across Different Approaches and Adaptation Strategies - Test Set

pivotal, with manual corrections significantly enhancing the performance of Translate-Train Monolingual models, underscoring the importance of precise alignment in cross-lingual tasks. GPT-4 achieved the highest scores, though with modest margins over specialized models, indicating the strong competitiveness of those models. Lastly, the results highlighted the critical impact of training data quality, as the Translate-Test approach outperformed Translate-Train due to the latter embedding errors from machine translation and annotation projection directly into the training data.

8 Qualitative Analysis on the Media Narratives on the Israeli War on Gaza

To assess the performance of the proposed model in detecting propaganda techniques, we conducted a qualitative analysis on the FIGNEWS subset (Section 3.3). These examples were selected to be annotated as propagandistic and to represent both narratives biased against Palestine and those biased against Israel. All annotated examples are in Appendix C.

The analysis of the examples reveals diverse strengths and shortcomings in the model's identification of propaganda techniques. Several examples showcase the model's ability to detect and label effectively, while others highlight areas for improvement in span detection and labeling accuracy.

The model performed strongly in identifying a variety of propaganda techniques, particularly in cases involving *Appeal to Fear*, *Appeal to Hypocrisy*, and *Loaded Language*. For instance, in Example 13:
"حذَّرنا إسرائيل من عواقب ملاحقة مسؤولين من حماس خارج فلسطين"
(We warned Israel about the consequences of pursuing Hamas officials outside Palestine) was accurately labeled as *Appeal to Fear*, as the phrase evokes concern about potential repercussions. Similarly, for *Appeal to Hypocrisy*, Example 7 includes:
"في الوقت الذي تحارب فيه إسرائيل إبادة شعب تُتهم هي بإبادة شعب"
(While Israel is fighting genocide, it is accused of genocide), which effectively exposes perceived inconsistencies in criticism. Another strong example of *Appeal to Hypocrisy* appears in Example 8:
"أين كرامة الإنسان، أين حقوق الإنسان، أين احترامه فالجواب هي إسرائيل"
(where human dignity is, where human rights are, and where respect is, the answer is Israel). These instances highlight the model's ability to identify rhetorical strategies that challenge or question the credibility of opponents.

The model also demonstrated proficiency in recognizing *Appeal to Time*, as seen in Example 6 with "المذبحة القادمة" (The next massacre) and "لن تنتهي الحرب قبل" (The war will not end before). Both spans emphasize urgency and the inevitability of action, aligning well with the intended technique. Additionally, the model's performance in labeling *Questioning the Reputation* was consistent across multiple exam-

ples. In Example 7, the span:
"أين كانت جنوب إفريقيا عندما قُتل وشُرد الملايين في سوريا واليمن على يد شركاء حماس"
(Where was South Africa when millions were killed and displaced in Syria and Yemen by Hamas's partners?) effectively critiques perceived hypocrisy, making the label appropriate. Similarly, in Example 8, "أين كرامة الإنسان" (Where is human dignity) and in Example 9,
"نتنياهو لا يفوت فرصة لالتقاط الصور لرفع أسهمه المتهاوية سياسياً" (Netanyahu never misses a chance to take pictures to boost his declining political ratings), were correctly identified as instances of questioning credibility.

The model's labeling of *Flag Waving* was another area of strength. For instance, in Example 7, the span:
"سنواصل الحفاظ على حقنا في الدفاع عن أنفسنا وتأمين مستقبلنا حتى النصر الكامل"
(We will continue to preserve our right to defend ourselves and secure our future until complete victory) was aptly labeled, as it appeals to patriotism and unity.

For *Exaggeration-Minimization*, Example 7 includes "الملايين" (Millions), while Example 5 includes "عملية عالية الجودة " (High-quality operation). Both spans are persuasive through their amplification of scale or quality, making the assigned labels fitting. Similarly, the *False Dilemma* technique is well-demonstrated in Example 12 with:
" لا تفاوض مع جيش الاحتلال بشأن تبادل الأسرى حتى انتهاء العدوان"
(No negotiations with the occupation army over prisoner exchange until the end of the aggression), which frames the situation as lacking alternatives. In Example 5, the span "وكل قادة حماس مصيرهم الموت " (All Hamas leaders are destined for death) similarly constructs a binary scenario, reinforcing the label's validity.

8.1 Limitations

Overprediction of Labels The model exhibited instances of overprediction, particularly for the *Loaded Language* label. For example, in Example 10, "تلّقي" (Receiving) was labeled as *Loaded Language*, despite being neutral. Similarly, in Example 13, "حذَّرنا" (We warned) was labeled as *Loaded Language*, though it does not carry an emotive or charged tone. Mislabeling was also seen in Example 12, where "قطاع" (Strip) was inaccurately labeled as *False Dilemma*, which does not align with the text's intent. In Example 6, "المخطوفين" (The captives) was labeled as *Name Calling*, but it is more descriptive than propagandistic.

Overly Broad or Irrelevant Spans The model demonstrated a tendency to select overly broad spans or include irrelevant elements within spans. For instance, in Example 13, the span "عواقب ملاحقة" (Consequences of pursuing) was labeled as *Loaded Language*, but only "consequences" carries the intended emotional charge, while "pursuing" is neutral. Similarly, in Example 8, "على دواعش حماس" (Over Hamas's Daesh) was labeled as *Questioning the Reputation*, but the inclusion of "على" (Over) extends the span unnecessarily.

Unidentified Propagandistic Content The model failed to identify certain propagandistic content. For Example 4, no spans were identified as propagandistic, yet the span
" إن القضاء على 'حماس' هو الطريقة الوحيدة لاستعادة الرهائن"
(Eliminating Hamas is the only way to retrieve the hostages) could be labeled as *False Dilemma* or *Appeal to Fear* due to its framing of a singular solution and invocation of threat.

In summary, the model demonstrates strong performance in recognizing clear techniques such as *Loaded Language*, *Name Calling*, and *Appeal to Fear*, but occasionally mislabels neutral phrases or includes extraneous content in spans. These findings underscore the importance of refining span selection and improving the accuracy of labels to handle nuanced cases effectively.

9 Conclusion

This work highlights the effectiveness of integrating argumentation features into propaganda detection models for Arabic texts. By combining claims, premises, and other argumentative elements with advanced NLP methodologies, we demonstrate consistent improvements over baseline models. Our analysis of Arabic media narratives reveals the model's ability to detect diverse propaganda techniques, offering valuable insights into rhetorical strategies in politically sensitive contexts.

While translation and annotation quality present challenges, the findings underscore the potential of this approach for fostering transparency in conflict-driven discourse. Future research should focus on refining annotation and translation methods. These advancements will contribute to building robust NLP tools capable of analyzing and mitigating the impact of propaganda in sensitive geopolitical contexts.

Acknowledgments

We acknowledge the assistance of the LT-Bridge Project (GA 952194) and DFKI for the use of their Virtual Laboratory.

References

Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2021. ARBERT & MARBERT: Deep bidirectional transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7088–7105, Online. Association for Computational Linguistics.

Sadaf Abdul Rauf, Huda Sarfraz, Saadia Nauman, Arooj Fatima, SadafZiafat SadafZiafat, Momina Ishfaq, Alishba Suboor, Hammad Afzal, and Seemab Latif. 2024. NLPColab at FigNews 2024 shared task: Challenges in bias and propaganda annotation for news media. In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 573–579, Bangkok, Thailand. Association for Computational Linguistics.

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the wanlp 2022 shared task on propaganda detection in arabic. *Preprint*, arXiv:2211.10057.

Khudejah Ali and Khawaja Zain ul abdin. 2021. Post-truth propaganda: heuristic processing of political fake news on facebook during the 2016 u.s. presidential election. *Journal of Applied Communication Research*, 49(1):109–128.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. AraBERT: Transformer-based model for Arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 9–15, Marseille, France. European Language Resource Association.

Mikel Artetxe, Vedanuj Goswami, Shruti Bhosale, Angela Fan, and Luke Zettlemoyer. 2023. Revisiting machine translation for cross-lingual classification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6489–6499, Singapore. Association for Computational Linguistics.

David Broniatowski, Daniel Kerchner, Fouzia Farooq, Xiaolei Huang, Amelia Jamison, Mark Dredze, and Sandra Crouse Quinn. 2020. The covid-19 social media infodemic reflects uncertainty and state-sponsored propaganda. *arXiv preprint*.

Alexander Chernyavskiy, Dmitry Ilvovsky, and Preslav Nakov. 2024. Unleashing the power of discourse-enhanced transformers for propaganda detection. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1452–1462, St. Julian's, Malta. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia. Association for Computational Linguistics.

Maram Hasanain, Firoj Alam, Hamdy Mubarak, Samir Abdaljalil, Wajdi Zaghouani, Preslav Nakov, Giovanni Da San Martino, and Abed Freihat. 2023. ArAIEval shared task: Persuasion techniques and disinformation detection in Arabic text. In *Proceedings of ArabicNLP 2023*, pages 483–493, Singapore (Hybrid). Association for Computational Linguistics.

Maram Hasanain, Md. Arid Hasan, Fatema Ahmed, Reem Suwaileh, Md. Rafiul Biswas, Wajdi Zaghouani, and Firoj Alam. 2024. Araieval shared task: Propagandistic techniques detection in unimodal and multimodal arabic content. *Preprint*, arXiv:2407.04247.

Ahmed Samir Hussein, Abu Bakr Soliman Mohammad, Mohamed Ibrahim, Laila Hesham Afify, and Samhaa R. El-Beltagy. 2022. NGU CNLP atWANLP 2022 shared task: Propaganda detection in Arabic. In *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 545–550, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Momtazul Labib, Samia Rahman, Hasan Murad, and Udoy Das. 2024. CUET_sstm at ArAIEval shared task: Unimodal (text) propagandistic technique detection using transformer-based model. In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 507–511, Bangkok, Thailand. Association for Computational Linguistics.

Salima Lamsiyah, Abdelkader El Mahdaouy, Hamza Alami, Ismail Berrada, and Christoph Schommer. 2023. UL & UM6P at ArAIEval shared task:

Transformer-based model for persuasion techniques and disinformation detection in Arabic. In *Proceedings of ArabicNLP 2023*, pages 558–564, Singapore (Hybrid). Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *Preprint*, arXiv:1907.11692.

Ana Laura Nettel and Georges Roque. 2012. Persuasive argumentation versus manipulation. *Argumentation*, 26(1):55–69.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multilingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

Md Riyadh and Sara Nabhani. 2024. Mela at ArAIEval shared task: Propagandistic techniques detection in Arabic with a multilingual approach. In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 478–482, Bangkok, Thailand. Association for Computational Linguistics.

Ali Safaya, Moutasem Abdullatif, and Deniz Yuret. 2020. KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 2054–2059, Barcelona (online). International Committee for Computational Linguistics.

Zien Sheikh Ali, Watheq Mansour, Tamer Elsayed, and Abdulaziz Al□Ali. 2021. AraFacts: The first large Arabic dataset of naturally occurring claims. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 231–236, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3):619–659.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation. *Preprint*, arXiv:2207.04672.

Yizhong Wang, Sujian Li, and Houfeng Wang. 2017. A two-stage parsing method for text-level discourse analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 184–188, Vancouver, Canada. Association for Computational Linguistics.

Wajdi Zaghouani, Mustafa Jarrar, Nizar Habash, Houda Bouamor, Imed Zitouni, Mona Diab, Samhaa R. El-Beltagy, and Muhammed AbuOdeh. 2024. The fignews shared task on news media narratives. *Preprint*, arXiv:2407.18147.

A Propaganda Techniques Definition

In this section, we provide the definitions of the propaganda techniques included in the dataset, as outlined in (Piskorski et al., 2023).

A.1 ATTACK ON REPUTATION

- **Name Calling-Labelling:** a form of argument in which loaded labels are directed at an individual, group, object or activity, typically in an insulting or demeaning way, but also using labels the target audience finds desirable.

- **Guilt by Association:** attacking the opponent or an activity by associating it with another group, activity, or concept that has sharp negative connotations for the target audience.

- **Doubt:** questioning the character or the personal attributes of someone or something in order to question their general credibility or quality.

- **Appeal to Hypocrisy:** the target of the technique is attacked based on their reputation by charging them with hypocrisy/inconsistency.

- **Questioning the Reputation:** the target is attacked by making strong negative claims about it, focusing specially on undermining its character and moral stature rather than relying on an argument about the topic.

A.2 JUSTIFICATION

- **Flag Waiving:** justifying an idea by exhaling the pride of a group or highlighting the benefits for that specific group.

- **Appeal to Authority:** a weight is given to an argument, an idea or information by simply stating that a particular entity considered as an authority is the source of the information.

- **Appeal to Popularity:** a weight is given to an argument or idea by justifying it on the basis that allegedly “everybody” (or the large majority) agrees with it or “nobody” disagrees with it.

- **Appeal to Values:** a weight is given to an idea by linking it to values seen by the target audience as positive.

- **Appeal to Fear-Prejudice:** promotes or rejects an idea through the repulsion or fear of the audience towards this idea.

A.3 DISTRACTION

- **Straw Man:** consists in making an impression of refuting an argument of the opponent’s proposition, whereas the real subject of the argument was not addressed or refuted, but instead was replaced with a false one.

- **Red Herring:** consists in diverting the attention of the audience from the main topic being discussed, by introducing another topic, which is irrelevant.

- **Whataboutism:** a technique that attempts to discredit an opponent’s position by charging them with hypocrisy without directly disproving their argument.

A.4 SIMPLIFICATION

- **Causal Oversimplification:** assuming a single cause or reason when there are actually multiple causes for an issue.
- **False Dilemma-No Choice:** a logical fallacy that presents only two options or sides when there are many options or sides. In extreme, the author tells the audience exactly what actions to take, eliminating any other possible choices.
- **Consequential Oversimplification:** is an assertion one is making of some “first” event/action leading to a domino-like chain of events that have some significant negative (positive) effects and consequences that appear to be ludicrous or unwarranted or with each step in the chain more and more improbable.

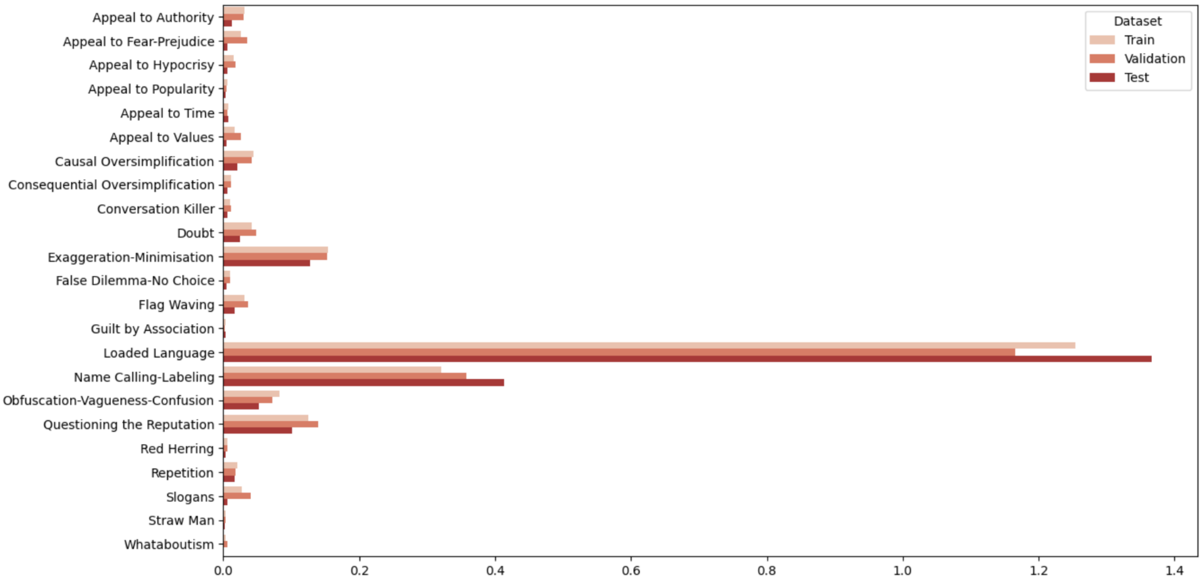
A.5 CALL

- **Slogans:** a brief and striking phrase, often acting like an emotional appeal, that may include labeling and stereotyping.
- **Conversation Killer:** words or phrases that discourage critical thought and meaningful discussion about a given topic.
- **Appeal to Time:** the argument is centered around the idea that time has come for a particular action.

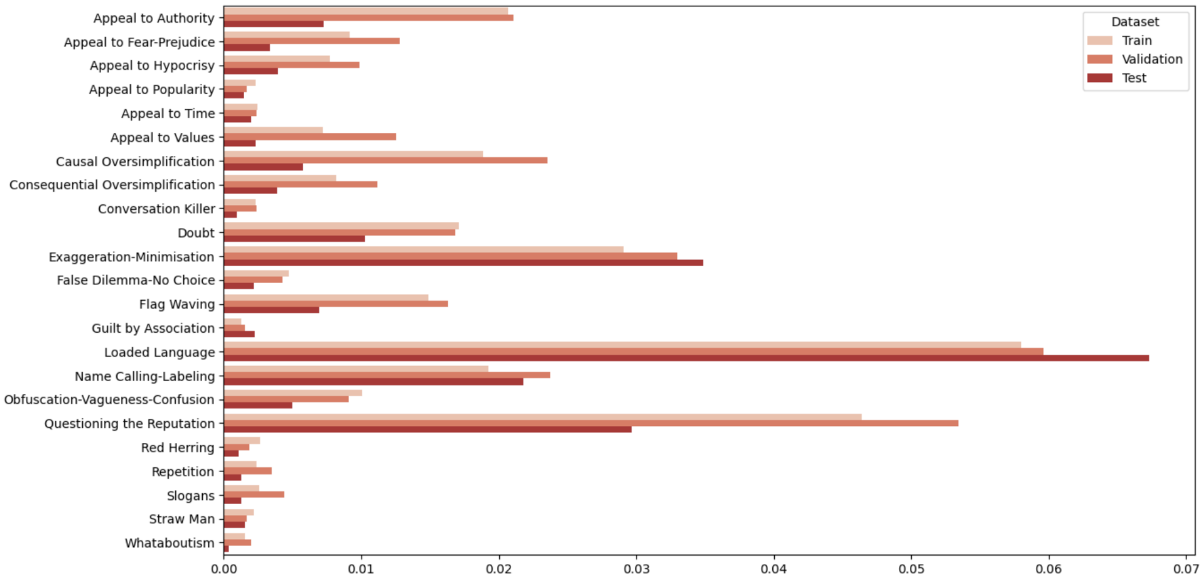
A.6 MANIPULATIVE WORDING

- **Loaded Language:** use of specific words and phrases with strong emotional implications (either positive or negative) to influence and convince the audience that an argument is valid.
- **Obfuscation-Vagueness-Confusion:** use of words that are deliberately not clear, vague, or ambiguous so that the audience may have its own interpretations.
- **Exaggeration-Minimisation:** consists of either representing something in an excessive manner or making something seem less important or smaller than it really is.
- **Repetition:** the speaker uses the same phrase repeatedly with the hope that the repetition will lead to persuading the audience.

B Distribution of propaganda techniques across the datasets



(a) Ratio of Labeled Documents to Total Documents



(b) Ratio of Labeled Tokens to Total Tokens

Figure 1: Propaganda Techniques Distribution: Ratios of Labeled Tokens and Documents

C Propaganda annotated examples of the narratives of the Israeli war on Gaza

Arabic	رسالة الى سكان حي الزيتون انتم تعرفون ان الحي مكتظ ومليئ بأوكار لحماس ولذلك وحفاظا على سلامتكم وعلى سلامة عائلاتكم واحبائكم توجهوا الى جنوب وادي غزة. كونوا متاكدين ان قادة حماس يتحصنون ويحاولون حماية أنفسهم
Translation	A message to the residents of the Zeitoun neighborhood: You know that the neighborhood is crowded and full of Hamas hideouts. Therefore, for your safety and the safety of your families and loved ones, move to the south of Wadi Gaza. Be assured that Hamas leaders are sheltering themselves and trying to protect themselves.
Labeled Spans	• مكتظ ("crowded") - Loaded Language • ومليئ ("full") - Loaded Language • متاكدين ("be assured") - Loaded Language • يتحصنون ("sheltering themselves") - Loaded Language

Table 4: Example 1 - Biased against Palestine

Arabic	الاتحاد الأوروبي يدرج رئيس المكتب السياسي لحركة حماس في قطاع غزة يحيى #السنوار على قائمة 'الإرهاب' المجلس الأوروبي: 'يندرج هذا القرار في إطار رد الاتحاد على التهديد الذي تشكله حماس وهجماتها الوحشية على #إسرائيل في السابع من أكتوبر' لتفاصيل أكثر
Translation	The European Union lists the head of Hamas' political bureau in the Gaza Strip, Yahya Sinwar, on the "terrorism" list. European Council: "This decision is part of the Union's response to the threat posed by Hamas and its brutal attacks on Israel on October 7." For more details.
Labeled Spans	• الإرهاب ("terrorism") - Loaded Language, Name Calling/Labeling • التهديد ("threat") - Loaded Language • وهجماتها الوحشية ("its brutal attacks") - Loaded Language

Table 5: Example 2 - Biased against Palestine

Arabic	الرئيس الأميركي جو_بايدن يعتبر أن استمرار المعارك في غزة قد يؤدي إلى تنفيذ أهداف حركة حماس بايدن: 'حماس شنت هجوماً إرهابياً لأنها لا تخشى شيئاً أكثر من أن يعيش الإسرائيليون والفلسطينيون جنباً إلى جنب في سلام' لتفاصيل أكثر
Translation	U.S. President Joe Biden considers that the continuation of battles in Gaza may lead to the achievement of Hamas' goals. Biden: "Hamas launched a terrorist attack because it fears nothing more than Israelis and Palestinians living side by side in peace." For more details.
Labeled Spans	• المعارك ("battles") - Loaded Language • شنت هجوماً إرهابياً ("launched a terrorist attack") - Loaded Language, Exaggeration/Minimization • تخشى شيئاً ("fears nothing") - Loaded Language • أن استمرار المعارك في غزة قد يؤدي إلى تنفيذ أهداف حركة حماس ("the continuation of battles in Gaza may lead to the achievement of Hamas' goals") - Causal Oversimplification • حماس شنت هجوماً إرهابياً لأنها ("Hamas launched a terrorist attack because") - Causal Oversimplification • حماس شنت هجوماً إرهابياً لأنها لا تخشى شيئاً أكثر من أن يعيش الإسرائيليون والفلسطينيون جنباً إلى جنب في سلام ("Hamas launched a terrorist attack because it fears nothing more than Israelis and Palestinians living side by side in peace") - Flag Waving

Table 6: Example 3 - Biased against Palestine

Arabic	وزير المالية الإسرائيلي بتسلئيل سموتريتش يقول إن القضاء على 'حماس' هو الطريقة الوحيدة لاستعادة الرهائن الأخبار حرب
Translation	The Israeli Minister of Finance, Bezalel Smotrich, says that eliminating 'Hamas' is the only way to retrieve the hostages.
Labeled Spans	None

Table 7: Example 4 - Biased against Palestine

Arabic	عاجل | 'يديعوت أحرونوت عن مسؤولين إسرائيليين': 'اغتيال العاروري عملية عالية الجودة وكل قادة حـ.ماس مصيرهم الموت'
Translation	Breaking: 'Yedioth Ahronoth citing Israeli officials': 'The assassination of al-Arouri is a high-quality operation, and all Hamas leaders are destined for death.'
Labeled Spans	• اغتيال (Assassination) - Loaded Language • عالية الجودة (High-quality) - Loaded Language • وكل قادة حـ.ماس مصيرهم الموت (All Hamas leaders are destined for death) - False Dilemma-No Choice • مسؤولين إسرائيليين (Israeli officials) - Obfuscation-Vagueness-Confusion • اغتيال (Assassination) - Name Calling-Labeling • عالية الجودة (High-quality) - Name Calling-Labeling • اغتيال العاروري عملية عالية الجودة وكل قادة حـ.ماس مصيرهم الموت (The assassination of al-Arouri is a high-quality operation, and all Hamas leaders are destined for death) - Appeal to Fear-Prejudice • اغتيال (Assassination) - Exaggeration-Minimisation • عملية عالية الجودة (High-quality operation) - Exaggeration-Minimisation

Table 8: Example 5 - Biased against Palestine

Arabic	عاجل | نتنياهو: - لن ننسى الفظائع التي وقعت في السابع من أكتوبر - نحن مصممون على تحقيق كل أهداف الحرب - لا بديل لنا عن النصر الساحق وإعادة مختطفينا في قطاع #غزة - لدينا الحق في الدفاع عن أنفسنا ولا أحد بإمكانه منعنا من ذلك - المذبحة القادمة بحق أبنائنا مسألة وقت لذلك يجب القضاء على حماس - وجهت الحكومة للقيام بمزيد من التفعيل لبرنامج صناعات دفاعية محلي لكي نعتمد على أنفسنا أكثر - أي تحقيقات يجب أن تتم بعد انتهاء الحرب - العلاقات مع #مصر تدار بشكل جيد ولكل بلد مصالحه التي يقلق بشأنها - لا أتراجع عن أي كلمة قلتها بخصوص #قطر - لن أتراجع عن أي مسار من مسارات الضغط على #حماس وقطر يمكنها القيام بهذا الضغط - قطر تستضيف قادة في حماس وبالتالي يمكنها ممارسة ضغط بخصوص المخطوفين - الموقف لا يزال على حاله بخصوص عدم إقامة مستوطنات في غزة - هدفنا القضاء على سلطة حماس ولا يمكن أن نسمح ببقاء قوات مسلحة في غزة ولن تنتهي الحرب قبل إكمال المهمة - محكمة العدل الدولية لم تجبرنا على إنهاء الحرب #حرب
Translation	Breaking: Netanyahu: - We will not forget the atrocities that occurred on October 7 - We are determined to achieve all the goals of the war - There is no alternative to decisive victory and the return of our captives in the Gaza Strip - We have the right to defend ourselves, and no one can prevent us from doing so - The next massacre against our children is a matter of time; therefore, Hamas must be eliminated - The government has been directed to further activate a local defense industries program to rely more on ourselves - Any investigations should take place after the war - Relations with Egypt are well-managed, and every country has its own interests to worry about - I do not back down from anything I said about Qatar - I will not back down from any pressure path on Hamas, and Qatar can exert such pressure - Qatar hosts Hamas leaders and can therefore exert pressure regarding the captives - The stance remains unchanged regarding the non-establishment of settlements in Gaza - Our goal is to eliminate Hamas authority, and we cannot allow armed forces to remain in Gaza; the war will not end before completing the mission - The International Court of Justice has not forced us to end the war.
Labeled Spans	• الفظائع (Atrocities) - Loaded Language • النصر الساحق (Decisive victory) - Loaded Language • المذبحة (Massacre) - Loaded Language • يقلق (Worried) - Loaded Language • الضغط (Pressure) - Loaded Language • المخطوفين (The captives) - Name Calling-Labeling • العلاقات (Relations) - Doubt • المذبحة القادمة (The next massacre) - Appeal to Time • مسألة وقت لذلك يجب القضاء على حماس (A matter of time; therefore, Hamas must be eliminated) - Appeal to Time • تنتهي الحرب قبل (Before the war ends) - Appeal to Time

Table 9: Example 6 - Biased against Palestine

Arabic	#عاجل | #نتنياهو: - في الوقت الذي تحارب فيه إسرائيل إبادة شعب تتهم هي بإبادة شعب - رأينا اليوم عالما مقلوبا رأسا على عقب ونحن نحارب الإرهابيين والأكاذيب - صراخ نفاق جنوب إفريقيا يصل إلى السماء - أين كانت جنوب إفريقيا عندما قتل وشرد الملايين في #سوريا واليمن على يد شركاء حماس - سنواصل الحفاظ على حقنا في الدفاع عن أنفسنا وتأمين مستقبلنا حتى النصر الكامل #حرب_غزة
Translation	#Breaking | #Netanyahu: - While Israel is fighting genocide, it is accused of genocide - Today we saw an upside-down world as we fight terrorists and lies - The hypocritical cries from South Africa reach the heavens - Where was South Africa when millions were killed and displaced in #Syria and Yemen by Hamas's partners? - We will continue to preserve our right to defend ourselves and secure our future until complete victory.
Labeled Spans	• إبادة شعب (Genocide) - Loaded Language • بإبادة شعب (Accused of genocide) - Loaded Language • الإرهابيين والأكاذيب (Terrorists and lies) - Loaded Language • صراخ نفاق (Hypocritical cries) - Loaded Language • وشرد الملايين (Displaced millions) - Loaded Language • في الوقت الذي تحارب فيه إسرائيل إبادة شعب تتهم هي بإبادة شعب - رأينا اليوم عالما مقلوبا رأسا على عقب ونحن نحارب الإرهابيين والأكاذيب - صراخ نفاق جنوب إفريقيا يصل إلى السماء - أين كانت جنوب إفريقيا عندما قتل وشرد الملايين في #سوريا واليمن على يد شركاء حماس (Where was South Africa when millions were killed and displaced in Syria and Yemen by Hamas's partners?) - Questioning the Reputation • في الوقت الذي تحارب فيه إسرائيل إبادة شعب تتهم هي بإبادة شعب - رأينا (While Israel fights genocide) - Appeal to Hypocrisy • مقلوبا (Upside-down) - Appeal to Hypocrisy • الإرهابيين (Terrorists) - Name Calling-Labeling • شركاء حماس (Hamas's partners) - Name Calling-Labeling • سنواصل الحفاظ على حقنا في الدفاع عن أنفسنا وتأمين مستقبلنا حتى النصر الكامل (We will continue to preserve our right to defend ourselves and secure our future until complete victory) - Flag Waving • أين كانت جنوب إفريقيا عندما قتل وشرد (Where was South Africa when killed and displaced) - Doubt • في #سوريا واليمن على يد شركاء حماس (In Syria and Yemen by Hamas's partners) - Doubt • الملايين (Millions) - Exaggeration-Minimisation

Table 10: Example 7 - Biased against Palestine

Arabic	سألني صديقي الجزائري من جديد: يا أفيخاي ما سر تفوق شعب إسرائيل على دواعش حماس ليس فقط عسكريًا. فأجابته: فإن سألت أين كرامة الإنسان، أين حقوق الإنسان أين احترامه فالجواب هي إسرائيل. وأشكر الله على كوني يهوديا وصهيونيا لإني منهما تعلمت ماذا يعني الحياة وماذا تعني الإنسانية.
Translation	My Algerian friend asked me again: Oh Avichai, what is the secret of Israel's superiority over Hamas's Daesh, not only militarily? I replied: If you ask where human dignity is, where human rights are, and where respect is, the answer is Israel. I thank God for being Jewish and Zionist because from them I learned what life and humanity mean.
Labeled Spans	• تفوق شعب (Superiority of a people) - Questioning the Reputation • على دواعش حماس (Over Hamas's Daesh) - Questioning the Reputation • أين كرامة الإنسان (Where is human dignity) - Questioning the Reputation • أين حقوق الإنسان أين احترامه فالجواب هي إسرائيل (Where are human rights and respect? The answer is Israel) - Questioning the Reputation • سألت (Asked) - Appeal to Hypocrisy • أين كرامة الإنسان، أين حقوق الإنسان أين احترامه فالجواب هي إسرائيل (Where is human dignity, where are human rights and respect? The answer is Israel) - Appeal to Hypocrisy • دواعش حماس (Hamas's Daesh) - Name Calling-Labeling • يهوديا وصهيونيا (Jewish and Zionist) - Name Calling-Labeling • سر تفوق شعب إسرائيل على دواعش حماس (The secret of Israel's superiority over Hamas's Daesh) - Doubt

Table 11: Example 8 - Biased against Palestine

Arabic	رئيس الوزراء الإسرائيلي يزور #غزة في 'وقت الهدنة' مع #حماس.. ورئيس منتدى الشرق الأوسط للدراسات السياسية والاستراتيجية: '#نتنياهو لا يفوت فرصة لالتقاط الصور.. لرفع أسهمه المتهاوية سياسياً' #فلسطين #إسرائيل #الحدث
Translation	The Israeli Prime Minister visits Gaza during the "time of the truce" with Hamas. The President of the Middle East Forum for Political and Strategic Studies says: "Netanyahu never misses a chance to take pictures to boost his declining political ratings."
Labeled Spans	• الهدنة (Truce) - Loaded Language • المتهاوية (Declining) - Loaded Language • #نتنياهو لا يفوت فرصة لالتقاط الصور.. لرفع أسهمه المتهاوية سياسياً (Netanyahu never misses a chance to take pictures to boost his declining political ratings) - Questioning the Reputation • الهدنة (Truce) - Name Calling-Labeling • الهدنة (Truce) - Appeal to Time

Table 12: Example 9 - Biased against Israel

Arabic	لحظة تلقّي والد الأسير المحرر بصفقة وفاء الأحرار والناطق باسم حركة حماس عن مدينة #القدس محمد حمادة نبأ استشهاد نجله المُبعد إلى #غزة من بلدة صور باهر ' #حرب_غزة #فيديو
Translation	The moment the father of the released prisoner under the "Wafa al-Ahrar" deal and spokesman for the Hamas movement in Jerusalem, Muhammad Hamada, received the news of the martyrdom of his son, who was displaced to Gaza from the town of Sur Baher.
Labeled Spans	• تلقّي (Receiving) - Loaded Language • المحرر (Released) - Loaded Language • استشهاد (Martyrdom) - Loaded Language • الأسير المحرر (Released Prisoner) - Name Calling-Labeling • وفاء الأحرار (Wafa al-Ahrar) - Name Calling-Labeling

Table 13: Example 10 - Biased against Israel

Arabic	مقال في صحيفة 'لوموند' الفرنسية جاء فيه أن مشاعر التعاطف مع ضحايا هجوم حماس عبر العالم تحوّلت بعد الهجوم على غزة نحو المدنيين الفلسطينيين بسبب معاناتهم.. أبرز ما ورد في الصحافة الدولية بشأن الحرب الإسرائيلية على قطاع غزة #حرب_غزة #الأخبار
Translation	An article in the French newspaper "Le Monde" stated that feelings of sympathy for the victims of Hamas's attack worldwide shifted after the attack on Gaza toward Palestinian civilians due to their suffering. Highlights from international press coverage of the Israeli war on Gaza.
Labeled Spans	• التعاطف (Sympathy) - Loaded Language • ضحايا هجوم (Victims of Attack) - Loaded Language • تحوّلت (Shifted) - Loaded Language • الهجوم (Attack) - Loaded Language • معاناتهم (Their Suffering) - Loaded Language • هجوم (Attack) - Name Calling-Labeling • عبر العالم (Worldwide) - Exaggeration-Minimisation

Table 14: Example 11 - Biased against Israel

Arabic	القيادي في حماس صالح العاروري في آخر لقاء تلفزيوني على شاشة #الجزيرة قبل استشهاده: لا تفاوض مع جيش الاحتلال بشأن تبادل الأسرى حتى انتهاء العدوان على قطاع #غزة #حرب_غزة #الأخبار
Translation	Hamas leader Saleh Al-Arouri in his last televised interview on Al Jazeera before his martyrdom: No negotiations with the occupation army over prisoner exchange until the end of the aggression on the Gaza Strip.
Labeled Spans	• استشهاده (Martyrdom) - Loaded Language • العدوان (Aggression) - Loaded Language • لا تفاوض مع جيش الاحتلال بشأن تبادل الأسرى حتى انتهاء العدوان (No negotiations with the occupation army over prisoner exchange until the end of the aggression) - False Dilemma-No Choice • قطاع (Sector/Strip) - False Dilemma-No Choice • جيش الاحتلال (Occupation Army) - Name Calling-Labeling • العدوان (Aggression) - Name Calling-Labeling

Table 15: Example 12 - Biased against Israel

Arabic	عاجل | رويترز عن مسؤول بالمخابرات التركية: 'حذرنا إسرائيل من عواقب ملاحقة مسؤولين من حماس خارج فلسطين بما فيها تركيا'
Translation	Breaking | Reuters quoting a Turkish intelligence official: ’We warned Israel about the consequences of pursuing Hamas officials outside Palestine, including in Turkey.’
Labeled Spans	• حذرنا (We warned) - Loaded Language • عواقب ملاحقة (Consequences of pursuit) - Loaded Language • حذرنا إسرائيل من عواقب ملاحقة مسؤولين من حماس خارج فلسطين (We warned Israel about the consequences of pursuing Hamas officials outside Palestine) - Appeal to Fear-Prejudice

Table 16: Example 13 - Biased against Israel

Arabic	ذكرت وسائل إعلام تابعة لحركة #حماس أن أكثر من 30 شخصا قتلوا وأصيب العشرات في قصف إسرائيلي لمخيم #جباليا في شمال.
Translation	Media affiliated with the Hamas movement reported that more than 30 people were killed and dozens injured in an Israeli bombing of Jabalia camp in the north.
Labeled Spans	• قصف (Bombing) - Loaded Language

Table 17: Example 14 - Biased against Israel

Arabic	هنية: اغتيال العاروري ورفاقه عملٌ إرهابيّ مكتمل الأركان وحماس لن تُهزَم
Translation	Haniyeh: The assassination of Al-Arouri and his companions is a fully-fledged terrorist act, and Hamas will not be defeated.
Labeled Spans	• اغتيال (Assassination) - Loaded Language • عملٌ إرهابيّ (Terrorist Act) - Loaded Language • العاروري ورفاقه (Al-Arouri and his companions) - Questioning the Reputation • الأركان (Fully-fledged) - Questioning the Reputation • لن تُهزَم (Will not be defeated) - Questioning the Reputation • مكتمل الأركان (Fully-fledged) - Obfuscation-Vagueness-Confusion • اغتيال العاروري (Assassination of Al-Arouri) - Name Calling-Labeling • عملٌ إرهابيّ (Terrorist Act) - Name Calling-Labeling • عملٌ إرهابيّ مكتمل الأركان وحماس لن تُهزَم (Fully-fledged terrorist act, and Hamas will not be defeated) - Exaggeration-Minimisation • هنية (Haniyeh) - Appeal to Authority • عملٌ إرهابيّ مكتمل الأركان (Fully-fledged terrorist act) - Appeal to Authority

Table 18: Example 15 - Biased against Israel

Arabic	ترحيب عربي بتدابير محكمة العدل الدولية بشأن منع الإبادة الجماعية في غزة، و'الصحة العالمية' ترفض اتهامات إسرائيلية بـ'التواطؤ مع حماس' تعرَّف أبرز أخبار اليوم
Translation	Arab approval of the measures by the International Court of Justice regarding preventing genocide in Gaza, and 'WHO' rejects Israeli accusations of 'collaboration with Hamas.' Discover the top #news of the day.
Labeled Spans	• الإبادة الجماعية (Genocide) - Loaded Language • اتهامات (Accusations) - Loaded Language • التواطؤ (Collaboration) - Loaded Language • 'الصحة العالمية' ترفض اتهامات إسرائيلية بـ'التواطؤ مع حماس' ('WHO rejects Israeli accusations of collaboration with Hamas') - Questioning the Reputation • 'التواطؤ مع حماس' ('Collaboration with Hamas') - Name Calling-Labeling

Table 19: Example 16 - Biased against Israel

Arabic	حركة حماس : تم الاتفاق مع الأشقاء في قطر ومصر على تمديد الهدنة الإنسانية المؤقتة لمدة يومين إضافيين بنفس شروط الهدنة السابقة.
Translation	Hamas Movement: Agreement was reached with the brothers in Qatar and Egypt to extend the temporary humanitarian truce for an additional two days under the same terms as the previous truce.
Labeled Spans	• الهدنة الإنسانية المؤقتة (Temporary Humanitarian Truce) - Loaded Language • الأشقاء (Brothers) - Name Calling-Labeling • الإنسانية (Humanitarian) - Name Calling-Labeling

Table 20: Example 17 - Biased against Israel