Sahara Pioneers at FIGNEWS 2024 Shared Task: Data Annotation Guidelines for Propaganda Detection in News Items

Marwa N. Solla, Hassan Ali Ebrahem, Alya Muftah Issa,
Harmain Harmain, Abdusalam F. A. Nwesri
College of Information Technology
University of Tripoli, Libya.

Abstract

In today's digital age, the spread of propaganda through news channels has become a pressing concern. To address this issue, the research community has organized a shared task on detecting propaganda in news posts. This paper aims to present the work carried out at the University of Tripoli for the development and implementation of data annotation guidelines by a team of five annotators. The guidelines were used to annotate 2600 news articles. Each article is labeled as "propaganda", "Not propaganda", "Not Applicable", or "Not clear". The shared task results put our efforts in the third position among 6 participating teams in the consistency track.

1 Introduction

Propaganda is a powerful tool often used to influence people's opinions and beliefs, especially during times of war and political contention. In the digital age, the spread of propaganda through online news and social media has become a major concern. Identifying propaganda is challenging as differentiating propaganda from other types of persuasion techniques is a difficult task. This is where artificial intelligence (AI) can play a valuable role.

The AI research community has developed numerous tools and techniques to help detect propaganda in news articles and other online content. A major trend in the development of AI-based propaganda detection tools relies heavily on the use of advanced machine learning (ML) models where the availability of training data is crucial.

The study of (Barron-Cedeno et al., 2019b) focused on binary-class propaganda detection at the news article level. Their model evaluates the level of propaganda based on the presence of keywords, the style of writing, and legibility. They also experimented on TSHP-17 and QProp corpora, where for the TSHP-17 corpus, they binarized the labels: propaganda vs. non-propaganda. Rashkin et al. (2017) focused on detecting various classes at a document-level where propaganda was one of four classes in their annotated data: satire, hoax, trusted, and propaganda. They adopted an analytical method to the language of news media in the context of political fact-checking and fake news detection and the articles were collected from the English Gigaword corpus and from seven other unreliable news sources. They trained their model by using word n-gram representation with logistic regression and reported that the model performed well only on articles from sources that the system was trained on.

A more fine-grained propaganda study was proposed by Da San Martino et al. (2019b), who developed a corpus of news articles annotated with 18 propaganda methods which was used in two shared tasks: SemEval-2020 (Da San Martino et al., 2020a) and NLP4IF-2020 (Da San Martino et al., 2019a). Besides that, improved models such as the Prta system (Da San Martino et al., 2020c), were proposed to address the limitations of transformers (Chernyavskiy et al.,2021). The Prta system was used to conduct a study of COVID-19 misinformation and

Proceedings of The Second Arabic Natural Language Processing Conference, pages 601–608

associated propaganda techniques in Bulgaria (Nakov et al.,2021a) and Qatar (Nakov et al., 2021b).

This paper presents our work on designing propaganda detection guidelines for the participation in FIGNEWS 2024 (Zaghouani et al., 2024), a shared task on news media narratives and part of the Second Arabic Natural Language Processing conference (ArabicNLP 2024). This shared task is more akin to a research-focused datathon with a strong emphasis on the development of improved annotation guidelines for complex opinion data tasks.

The rest of the paper is organized as follows: Section 2 explains our methodology and annotation process. Section 3 provides some examples of the annotation process and in Section 4 we present the results as reported by the organizers of the shared task. In Section 5 we discuss how the guidelines were used in the annotation process. Section 6 gives some concluding remarks and future work.

2 Methodology

2.1 Team and Expertise

Our team consisted of five participants from the University of Tripoli, College of Information technology, each with a diverse background in information technology and computer sciences. Supervised by two senior experts, the team decided to participate in the propaganda detection track of FIGNEWS 2024. Their combined expertise allowed for writing concise and clear propaganda annotation guidelines (see appendix) and using these guidelines to annotate the dataset given by the shared task organizers. Furthermore, to simplify the annotation process, an in-house tool was developed. The tool allows a group of users to work on the dataset at the same time to speed up the process.

2.2 Development of Data Annotation Guidelines

The team used annotation categories defined in the shared task and developed a set of heuristic guidelines to establish clear criteria for propaganda detection. Frequent group discussions and iterations were conducted to ensure clarity and consistency in the guidelines.

2.3 Preparing the Dataset for Annotation

The dataset provided by the shared task organizers was used in three stages:

1- We selected an experimental set of 100 posts. This small set is used for the initial discussions and to familiarize the team with the task.
2- During writing the annotation guidelines we used related dataset generated by ChatGPT.
3- In the final stage we uploaded the whole dataset to our inhouse annotation tool and followed the guidelines to annotate it. The result of the annotation tool was exported back into the excel sheet.

2.4 Annotation Process

To provide the annotation team knowledge of the proposed guidelines, a thorough training session was conducted. The purpose of this training was to establish a collective comprehension of the propaganda detection process and to address any questions or concerns. Following that, a meeting was scheduled to collectively annotate a sample of news data and solicit the team assessment and feedback.

In order to settle any disputes among the annotators, the team leader actively participated in the process and organized discussion sessions to ensure that everyone is in agreement on each contested annotation. The team then started working on annotating the first and second batches' Inter Annotation Agreements (IAAs). It was split among annotators, along with the remaining first and second batches of the MAIN, in order to prevent duplicate annotations. The annotator was allowed to consult with the expert member on some cases where the guidelines are not clear or partially applicable. Lastly, to assess the overall quality and consistency of the annotations, the team leader examined random samples of the annotated dataset.

	Quantity	Quality				Centrality				Guidelines	
Team	Data Points	IAA Kappa	Acc	F1 Avg	F1 Prop*	Kappa	Acc	Macro F1 Avg	F1 Prop*	Document Score	weighted Document + IAA Kappa
NLPColab	**16,500**	**69.9**	87.3	73.1	92.1	21.5	53.3	**39.9**	61.6	**7**	**0.9375**
Sina	**13,200**	**65.3**	85.7	67.9	76.7	12.5	48.1	27.7	44.5	**3**	**0.6551**
The CyberEquity Lab	**4,200**	**33.5**	65.3	41.1	54.2	22.1	55	**37.4**	55.7	5	0.5521
Bias Bluff Busters	2,600	31.5	54.3	47.3	62.6	20.2	51.1	**37.6**	54.2	**7**	**0.6632**
Sahara Pioneers	2,600	27.9	49.1	46.3	50.5	18.7	48.7	**37.4**	56.3	7	0.6373
Narrative Navigators	2,200	12.8	54.5	54.6	63.1	19.4	52.7	37.1	60.7	8	0.5913

Table 1: Results as reported by shared Task organizer

3 Examples of Annotated News Items

Here we provide some examples of annotated news posts to showcase our annotation process. These examples highlight the identification of various propaganda techniques and provide an interpretation of the annotation results.

Identifying Propaganda:

"FREE AT LAST: The identities of the second group of 13 Israeli hostages freed from Hamas terrorists have been revealed. The group consists of women, teenagers and children, and many of them are from Kibbutz Be'eri, one of the kibbutzim that was devastated by Hamas. https://trib.al/8Inc6oq:=:https://www.foxnews.com/world/israel-hamas-war-identities-recently-released-israeli-hostages-revealed"
This news item was classified as Propaganda because of the detection of a propagandistic technique that uses emotional stimuli was employed to influence opinions and behaviors rather than relying on a logical explanation.

Identifying not-propaganda

“The name of an IDF casualty whose family was notified is attached: - Sgt. Gilad Rozenblit, 21 years old, Magnigar, a combat medic in the 52nd Battalion, 401st Brigade (forming the Iron Tracks), fell during an operational activity in the north Gaza Strip last night (Thursday)."

This news post is classified as Not-Propaganda, because it contains accurate, factual information and a balanced presentation of information.

Handling Unclear Cases

When a news post is ambiguous or lacks sufficient information to determine its propaganda status, annotators label it as Unclear. An example of this are news items such as:

"The Universal Declaration of Human Rights announced the reaction of the Palestinians #Gaza #Israel #terrorism #condemnation #Violence ".

Handling Not Related Cases

Annotators label a news post as "Not Related" if it is not relevant to the topic of the shared task such as:" PUBG Mobile explodes a surprise with exciting details that excite players."

4 Results

In the Propaganda subtask, there were 51 annotators across 6 teams, and they annotated together a total of 41,300 news posts. For each subtask, there are four evaluation tracks: Quantity, Quality, Consistency, and Guidelines. As shown on Table 1, in the Quantity track we annotated 2600 data points and that satisfied the shared task requirements. In the IAA Quality, the primary metric was IAA Kappa score which was calculated as the average of all pairwise Kappa scores between team annotators for each relevant IAA batch and subtask. Our Kappa score was 18.7 where the highest score in this track was 21.5, the lowest was 12.5, and the average score was 19. In the Consistency Track, teams compete based on the centrality of their annotation choices

compared to all other teams, and we came on the 3rd place. For the Guidelines track we satisfied all requirements and scored 7 out of 8.

5 Discussion

5.1 Evaluation of the Annotation Process

The guidelines we designed to propaganda subtask proved to be suitable for detecting propaganda in news posts. Our team came on 3rd place in the Consistency track, and this is a strong indicator that our guidelines were effective in helping annotators recognize clear instances of propaganda. The collaborative efforts of the team significantly contributed to the success of the annotation process. Regular team discussions and calibration sessions helped ensure a shared understanding of the guidelines, which further contributed to the overall consistency and reliability of the annotations. The fact that most of the annotators agreed with each other shows that the guidelines gave clear, usable criteria for identifying different propaganda techniques. Drawing on our research experience, we highlight the need of precise and unambiguous criteria to facilitate an efficient annotation process.

5.2 Limitations and Challenges

The main barriers were time constraints, and the ambiguity and subjectivity involved in defining propaganda. Our team's background is in computer science and IT; if we had included people with a background in data science, we could have produced better results. To enhance our set of guidelines, further research ought to involve a more varied set of annotators.

6 Conclusion

This research paper presented our work on designing a set of annotation guidelines for propaganda detection in news items. These guidelines were designed through an iterative process involving thorough discussions, and empirical testing. The proposed guidelines detail how to identify and label propaganda techniques in text and define clear, unambiguous categories. Furthermore, to address propaganda interpretation issues, illustrative examples were included into the guidelines. This method aids in guaranteeing that all annotators possess an identical comprehension of the criteria for annotation. The guidelines have shown that they significantly improve inter-annotator agreement, indicating annotation consistency. This improvement is essential for datasets for propaganda detection, machine learning model training, and evaluation. Future work will refine these guidelines based on feedback from the research community. Additionally, the team plans to incorporate these guidelines into automated annotation tools to optimize the process. In conclusion, we believe that our data annotation guidelines will help tp advance the systematic study of news media propaganda.

References

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

Anton Chernyavskiy, Dmitry Ilvovsky, and Preslav Nakov. 2021. Transformers: "*The end of history" for NLP? In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases,* ECMLPKDD'21.

Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020. SemEval-2020 task 11: Detection of propaganda techniques in news articles. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1377–1414, Barcelona (online). International Committee for Computational Linguistics.

Giovanni Da San Martino, Seunghak Yu, Alberto Barr´on-Cede˜no, Rostislav Petrov, and Preslav Nakov. 2019b. *Fine-grained analysis of propaganda in news article*s. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLPIJCNLP '19, pages 5636–5646, Hong Kong, China.

Giovanni Da San Martino, Shaden Shaar, Yifan Zhang, Seunghak Yu, Alberto Barr´on-Cedeno, and Preslav Nakov. 2020c. *Prta: A system to support the analysis of propaganda techniques in the news.* In Proceedings of the Annual Meeting of Association for Computational Linguistics, ACL '20, pages 287– 293.

Giovanni Da San Martino, Alberto Barron-Cedeno, and Preslav Nakov. 2019a. *Findings of the NLP4IF-2019 shared task on fine-grained propaganda detection. In Proceedings of the 2nd Workshop on NLP for Internet* Freedom (NLP4IF): Censorship,

Disinformation, and Propaganda, NLP4IF ’19, pages 162–170, Hong Kong, China.

Hasanain, M., Ahmed, F., & Alam, F. (2024). *Can GPT-4 Identify Propaganda? Annotation and Detection of Propaganda Spans in News Articles.* arXiv preprint arXiv:2402.17478.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. *Truth of varying shades: Analyzing language in fake news and political fact-checking*. In Proceedings of the 2017 conference on empirical methods in natural language processing, pages 2931–2937

Pankaj Gupta, Khushbu Saxena, Usama Yaseen, Thomas Runkler, and Hinrich Schutze. 2019. *Neural Architectures for Fine-Grained Propaganda Detection in News*, https://doi.org/10.48550/arXiv.1909.06162

Preslav Nakov, Firoj Alam, Shaden Shaar, Giovanni Da San Martino, and Yifan Zhang. 2021b. *A second pandemic? Analysis of fake news about COVID- 19 vaccines in Qatar*. In Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP ’21.

Preslav Nakov, Firoj Alam, Shaden Shaar, Giovanni Da San Martino, and Yifan Zhang. 2021a. COVID-19 in *Bulgarian social media: Factuality, harmfulness, propaganda, and framing*. In Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP ’21.

Wajdi Zaghouani, Mustafa Jarrar, Nizar Habash, Houda Bouamor, Imed Zitouni, Mona Diab, Samhaa R. El- Beltagy, and Muhammed Raed AbuOdeh, editors. 2024. *The FIGNEWS Shared Task on News Media Narratives*. Association for Computational Linguistics, Bangkok, Thailand. [Online].available: https://sites.google.com/view/fignews/home .

Appendix I: Guidelines

Tripoli University
Team name: Sahara Pioneers

Guidelines for Subtask 2: Propaganda Identification

Task Description

The purpose of this annotation task is to identify and classify news posts based on their potential to contain propaganda. The task is to annotate news posts with one of four labels: Propaganda, Not Propaganda, Unclear, or Not Applicable.

Objectives

The guidelines are intended to fulfill the following objectives:

1. Promote consistency in the data classification among annotators.
2. Improves the labeling accuracy by reducing the misclassed data.
3. Provide a reference point for review and feedback, enhancing the quality of classified data.
4. Provide a thorough resource to help new annotators grasp the nuances of each category, allowing for effective classification.
5. Establish a shared language of communication to handle any concerns that may occur during the process.

Labels:

1. **Propaganda**: This label should be assigned to news postings that intentionally manipulate information to support a specific goal or position (for example, those that give fabricated facts or utilize emotional language to manipulate readers).
2. **Not Propaganda**: This label should be assigned to news posts that do not display any obvious propaganda elements.
3. **Unclear**: This label should be assigned to news posts where it is difficult to determine whether they contain propaganda or not.
4. **Not Applicable**: This label should be assigned to news posts that do not fit the scope of the annotation task.

Labeling Process:

The guidelines are applied to assess the presence or absence of propaganda by considering various factors and characteristics of the news posts.

1. Identifying Propaganda:
    a. Look for signs of deliberate misinformation in the news post. Pay attention to the propaganda techniques given below.
    b. Consider the presence of exaggerated claims, emotional manipulation, or one-sided presentation of information.
    c. Assess whether the news post aims to manipulate public opinion by intentionally spreading false or misleading information.
    d. Consider the credibility and reputation of the news source when evaluating the likelihood of propaganda.
2. Identifying Not-Propaganda:
    a. Focus on the factual accuracy and balanced presentation of information in the news post.
    b. Look for evidence-based reporting and citations of reliable sources.
    c. Assess whether the news post adheres to journalistic standards of fairness, objectivity, and transparency.
    d. Evaluate whether the post provides verifiable facts and supports claims with evidence.
3. Handling Unclear Cases
    a. When a news post is ambiguous or lacks sufficient information to determine its propaganda status, annotators label it as "Unclear."
    b. Consider factors such as insufficient context, conflicting information, or difficulty discerning the intent or bias of the post.
    c. Avoid making assumptions or speculation and base their judgment on the available information.

4. Handling Not Related Cases
    a. Annotators label a news post as "Not Related" if it is not relevant to the topic of propaganda or does not contain any news content.
    b. They consider whether the post is an advertisement, personal opinion, or unrelated content that doesn't contribute to the discussion of propaganda.

Propaganda Techniques

The following are some of the propaganda techniques used in news articles and posts.

1. Name-calling: Name-calling is the use of derogatory or sarcastic language to establish a negative perception of a person, group, or idea without supporting proof or logic. Example: "The actions of Hamas rebels are nothing short of terrorist attacks."
2. Stereotyping: the practice of making broad generalizations or assumptions about a certain group based on incomplete or biased information, which frequently results in inaccurate perspectives. Example: "All Palestinians are supporting terrorist organizations."
3. Loaded Terms: Utilizing emotionally charged or biased language to influence perception and elicit a strong emotional response from the audience. Example: " Hamas's brutal actions are the root cause of the conflict"
4. Appeal to Authority: a tactic that involves referencing anonymous experts or figures of authority to increase credibility and persuade others to adopt a particular viewpoint or do a specific action. Example: "Renowned scholars agree that Israel's policies constitute apartheid. Shouldn't we listen to their expertise?"
5. Emotional Appeals: The use of emotional triggers such as empathy, compassion, or guilt to influence opinions and behaviors rather than depending on logical explanation. Examples: "How can we stand by while innocent Israeli children suffer under Hamas daily attacks."
6. Exaggeration: Magnifying specific qualities of a situation, person, or group to stress their importance or influence while potentially misrepresenting the reality. Example: "Hamas is the most dangerous organization, and our military are doing their job right."
7. Dehumanization: The portrayal of individuals or groups as less than humans, typically using a charged language to justify abuse, discrimination, or violence against them. Example: "Why should we care about the rights of terrorists who target innocent Israeli civilians?"
8. Lack of reference: omitting references, sources, or evidence to support claims or arguments, making it difficult for others to fact-check or verify the information provided. Example: "Surveys show that the majority of Israelis support the blockade on Gaza, proving it is justified."
9. Fear-Mongering: Exaggerating fear or panic among the public through misleading information, often without explicitly promoting a specific action or viewpoint. Example: "Any criticism of Israel should be labeled as anti-Semitic and silenced?"
10. Distorted Statistics: Manipulating or presenting statistics in a misleading or selective manner to support a particular narrative or agenda, often by omitting relevant information or misinterpreting data. Example: "Statistics show that the majority of Palestinians support the use of violence against Israelis"

Quality Check Procedure:

1. Conduct regular meetings with annotators to address questions and provide clarifications.
2. Randomly sample a portion of annotated posts for review by senior annotators or experts.
3. Compare the annotations of the sampled posts to ensure consistency and accuracy.
4. Provide feedback to annotators based on the review results to improve their performance.

Handling Ambiguity and Ensuring Consistency:

1. Encourage annotators to discuss ambiguous cases with their peers or team leader for consensus.
2. Maintain a shared document or forum where annotators can seek clarification on specific posts or cases.
3. Provide clear guidelines on the criteria for assigning each label to minimize confusion and inconsistency.
4. Conduct regular calibration exercises to ensure annotators interpret the guidelines consistently.
5. Ethical Considerations:
    a) Emphasize the importance of impartiality and fairness in the annotation process.
    b) Maintain strict confidentiality and data security protocols to protect the privacy of news posts and annotators.
    c) Avoid bias by selecting a diverse group of annotators with varied backgrounds and perspectives.
    d) Regularly assess and address any potential biases or conflicts of interest among annotators.