



https://araieval.gitlab.io

ArAIEval Shared Task: Propagandistic Techniques Detection in Unimodal and Multimodal Arabic Content



Northwestern University

Maram Hasanain, Md. Arid Hasan, Fatema Ahmed, Reem Suwaileh,
Md. Rafiul Biswas, Wajdi Zaghouani, Firoj Alam

1 Unimodal (Text) Propagandistic Technique Detection

Propaganda is a communication tool that is **deliberately** used to **influence** audience **towards a specific goal**

Task Definition: Given a multigenre text snippet, detect the propaganda techniques used in the text together with the exact span(s) in which each propaganda technique appears.

أضاف: "وبالتوازي مع الأجواء التفاؤلية التي تبتئها مصادرهم يتولون رمي الإشاعات بأن عملية التأليف انتهت، وبانت مسألة ساعات، وإيهام الآخرين بأنهم يريدون تشكيل الحكومة وغيرهم يريد التعتيل، تماماً كما فعلوا في الحكومات السابقة، وبالأخص في وزارة الطاقة التي استمرت لسنوات بسببهم وتكثرت تآسين الكهرباء 24 ساعة على 24، فإن الكهرباء وقد بلغ الهدر في هذا القطاع 55 في المئة من مالية الدولة، وها هم مستمرين في التعتيل ليقولوا للناس "فلان"، ولهذا السبب كانوا يخشون من الحريري".

Translation: He added: "In parallel with the optimistic atmosphere that their sources are spreading, they are spreading rumors that the formation process has ended, and that it's only a matter of hours. They deceive others into believing that they want to form the government while others want to obstruct it just as they did with previous governments, particularly in the Ministry of Energy, which they held for years and promised to provide electricity 24/7. So where is the electricity, given that the waste in this sector has reached 55% of the state's finances, they continue to obstruct to tell people 'they didn't let us' and that's why they were afraid of Hariri."

Techniques: Appeal to Authority Smears Appeal to Fear Bandwagon

Example annotated paragraph

Data Collection:

- Tweets** collected from different accounts of Arabic news sources (Hasanain et al., 2023) + tweets relevant to war on Gaza.
- News paragraphs** selected from news articles (Hasanain et al. 2023)
 - a. AraFacts (Ali et al., 2021)
 - b. in-house news articles collection

Annotation:

- Phase 1:** Individual annotators annotate the dataset
- Phase 2:** Consolidation is done with expert annotators to resolve disagreement and ensure quality.

Dataset size

1.5K tweets + 7.5K paragraphs

2 Multimodal Propagandistic Memes Classification

Task Definition

- Subtask 2A:** Given a text extracted from a meme, categorize whether it is propagandistic or not.
- Subtask 2B:** Given a meme (text overlayed image), detect whether the content is propagandistic.
- Subtask 2C:** Given multimodal content (text extracted from meme and the meme itself), detect whether the content is propagandistic.

Translation: when you think about all the opportunities you have for success but then you remember that you live in an Arab country

لما تتخيل فرص النجاح اللي قدامك بس بعدين تتذكر انك عايش بدوله عربييه

a) Propagandistic

Data Collection:

- Memos** collected from public groups on social media platform (e.g., Facebook, Twitter (X), Instagram, etc)
- Texts from the memes were extracted using an off-the-shelf OCR.

Annotation:

- Three annotators per meme
- Majority voting for final labels

Dataset size

3K Memos

Results

TASK 1

Team	Rank	Micro F1
CUET_sstm	1	0.2995
Mela	2	0.2833
MemeMind	3	0.2774
Nullpointer	4	0.2541
SussexAI	5	0.1228
SemanticCuetSync	6	0.0783
Baseline		0.0151

TASK 2A

Team	Rank	Macro F1
AlexUNLP-MZ	1	0.787
CLTL	2	0.779
MemeMind	3	0.746
DLRG	4	0.739
One_by_zero	5	0.674
Z-Index	6	0.633
Baseline		0.453

TASK 2B

Team	Rank	Macro F1
CLTL	1	0.711
MemeMind	2	0.664
AlexUNLP-MZ	3	0.659
Baseline		0.475

TASK 2C

Team	Rank	Macro F1
AlexUNLP-MZ	1	0.805
ASOS	2	0.798
CLTL	3	0.798
MemeMind	4	0.797
Team Engima	5	0.753
MODOS	6	0.729
Z-Index	7	0.712
Baseline		0.493

Evaluation Setup

- Development phase:** released train and development subsets, and participants submitted runs on the **development subsets**.
- Test phase:** participants submitted runs on the official **test subsets**.

Participation

Total (test phase): 14 teams
Task 1: 6 teams Task 2: 9 teams

Approaches

- Task 1:** fine-tuning Arabic transformer models like AraBERT is the most common system architecture. Majority of the teams modeled the task as a token classification problem.
- Task 2:** fine-tuning transformer models such as MARBERT is the most popular architecture. As for the vision models ResNet was the most popular choice. For subtask C, multiple fusion techniques of text and image models were used.

Findings

- Task 1:**
 - Results demonstrate the difficulty of the task with the best system achieving a Micro F1 of 0.3.
- Task 2:**
 - As the subtasks were binary classification tasks, the participating systems generally had strong performance. The subtask 2B that is focused on the image modality proven to be the most challenging.

Summary and Future Work

Summary

- The shared task tackled both text and image modalities.
- Some systems reported challenges due to the skewed label distribution and attempted data augmentation approaches.
- Most systems fine-tuned transformer models.

Future work

- Extend the multimodal task to be at the span level
- Offer span level detection tasks

Acknowledgments

The work of M. Hasanain, F. Ahmed, R. Suwaileh, W. Zaghouani and F. Alam is partially supported by NPRP 14C-0916-210015 from the Qatar National Research Fund, which is a part of Qatar Research Development and Innovation Council (QRDI). The findings achieved herein are solely the responsibility of the authors.

