Exploring the Dataset Landscape for Automated Propaganda Detection: A Data-Centric Insight

Marco Usai[1], Davide Antonio Mura[1], Andrea Loddo*[1], Manuela Sanguinetti[1], Luca Zedda[1], Cecilia Di Ruberto[1] and Maurizio Atzori[1]

[1]Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, Cagliari, 09124, Italy

Abstract

The increasing spread of propaganda in digital media has intensified research efforts toward the development of automated detection systems. Central to this task is the availability and quality of annotated datasets, which directly impact model performance, generalizability, and real-world applicability. In this paper, we present a data-centric insight into the current landscape of datasets used for automated propaganda detection. We analyze a representative set of publicly available corpora with respect to key factors such as annotation schemes, label granularity, domain coverage, linguistic diversity, and class balance. This work aims to guide researchers toward more robust, inclusive, and scalable approaches to propaganda detection by emphasizing the foundational role of data quality and structure.

Keywords

Propaganda Detection, Span Identification, Dataset Benchmarking

1. Introduction

Propaganda is a deliberate and systematic form of communication that aims to influence the opinions, beliefs, and behaviours of a target audience. This influence is often exerted through the selection, omission, or distortion of information. The employment of sophisticated persuasive techniques by the sender is evident, with the utilisation of emotional appeals, oversimplification, and the use of stereotypes and slogans. The outcomes of this communication are primarily oriented towards the interests of the sender rather than those of the recipient.

Historically, propaganda has played a central role in a variety of contexts, from religion and politics to modern mass media and public relations systems. In the contemporary digital age, the capacity to discern propaganda is of paramount importance. Hyperconnectivity and the deluge of information have led to an escalation in the risk of manipulation through misinformation, fake news, and targeted messaging designed to polarise public opinion.

In such contexts, the ability to discern propaganda is crucial for promoting critical reflection, preserving the integrity of public discourse, and ensuring the proper functioning of democratic processes. This paper presents a mini-review and analysis of 10 research studies on propaganda detection, published between 2021 and 2025. For each work, an examination is conducted of the employed dataset, including its annotation scheme and characteristics, the methodological approach adopted to address the problem, and the results achieved. Due to limitations in available space, the decision has been made to select primarily works that address the problem using a specific annotation scheme, namely span identification.

This choice is motivated by the fact that span detection enables a more fine-grained and interpretable analysis of propaganda techniques. Unlike binary classification, which merely labels entire texts or documents as propagandistic or not, span detection allows researchers to pinpoint the exact textual

segments where propaganda occurs, along with the specific technique being used. This granularity is essential for understanding how propaganda is constructed and communicated within a text, facilitating more informative downstream applications such as fact-checking, media literacy education, and automated content moderation.

Recent advancements in AI have led to the development of data-centric AI (DCAI), in which data, rather than models, is recognized as the primary driver of robust and adaptable systems. As outlined by Malerba and Pasquadibisceglie [1], central to DCAI is the understanding that datasets are dynamic, evolving resources that require ongoing curation, enrichment, and validation. This principle is particularly critical for propaganda detection because the tactics, linguistic nuances, and dissemination strategies of propaganda actors are fluid and adapt quickly to sociopolitical contexts. Therefore, datasets for automated propaganda detection must be conceptualized as living resources that are continuously refined to maintain relevance and effectiveness in a rapidly changing landscape.

Our review of ten prominent propaganda detection corpora reveals a consistent pattern. Although annotation schemes have become more detailed and multilingual coverage has improved, most datasets remain largely static after release, with little evidence of continued maintenance or updates. The discrepancy between dataset stasis and task dynamism underscores the necessity of aligning future dataset development with the DCAI paradigm.

The structure of the paper is thus organized as follows: Section 2 outlines the criteria adopted for selecting the reviewed works and describes the research methodology, Section 3 provides an overview of the selected studies, while Section 4 offers a synthesis of key findings and discusses broader implications and conclusions.

2. Search Protocol

The primary criterion for selecting the literature reviewed in this survey was the relevance and contribution of existing datasets for automatic propaganda detection, with particular attention to those supporting span-level annotations. Rather than focusing solely on methodological innovations, the survey emphasizes corpora that have played a significant role in shaping recent research directions and enabling fine-grained analysis of propaganda in online content.

We thus conducted a literature search primarily focusing on Google Scholar and using a specific set of keywords, i.e., "computational propaganda detection", "social media manipulation detection", "automated propaganda detection", and "misinformation detection". These keywords were selected to encompass a wide range of approaches that address the detection of persuasive, manipulative, or misleading content on social media and other digital platforms. Priority was given to peer-reviewed papers that introduce documented and publicly available datasets, either accessible via direct download or obtainable upon request from the dataset curators. The selection focused in particular on resources with the following main characteristics:

- Span-level annotations, allowing for fine-grained identification of propagandistic content within texts;
- Rich annotation schemes capturing related phenomena, such as persuasion techniques, manipulative strategies, or framing categories.

These criteria aimed to ensure the inclusion of high-quality datasets that support both detailed analysis and replicability in propaganda detection research.

Due to the very focused search scope, we found 10 datasets, all released between 2019 and early 2025. Despite their recent release, many of them have been widely used as benchmark datasets to develop or test computational approaches, thus highlighting the wide interest in this topic and the related tasks that these resources aim to support. To prove this, Table 1 summarizes the results of this search, along with the publication year and number of citations as available on Google Scholar.

While the next section briefly outlines the datasets found in our search, Section 4 aims to broadly discuss the main findings in terms of specific key factors, particularly oriented to a more data-centric perspective.

Table 1
Overview of key datasets for propaganda detection, sorted by year of release. The table reports the dataset name, year of publication, and number of citations as of mid-2025 (according to Google Scholar, when available).

Dataset	Year	# cit.
PTC	2019	433
PTC-SemEval 2020	2020	255
WANLP	2022	77
SemEval 2023 Task 3	2023	123
BanMANI	2023	1
Salman et al.	2023	15
ArAIEval	2024	3
ArMPro	2024	29
ZenPropaganda	2024	-
ManiTweet	2025	8

3. Datasets Overview

This section summarizes the retrieved information about the identified datasets, also providing Table 2 as a reference table.

Propaganda Techniques Corpus (PTC) [2]   PTC is a corpus designed for fine-grained propaganda detection in news articles. It contains 451 English news articles annotated at the span level, where each propagandistic fragment is marked and labeled with one of 18 propaganda techniques. The annotations include precise character offsets for each span. The articles were sourced from both mainstream and suspicious outlets, and the annotations were conducted by trained experts.

PTC-SemEval 2020 [3]   The dataset was proposed for the SemEval-2020 Task 11 on Propaganda Detection, and it precisely builds upon the theoretical framework devised initially for the seminal PTC dataset. It thus focuses on detecting propaganda techniques in English news articles, and it was designed to address two subtasks: span identification (detecting specific text spans containing propaganda) and technique classification (assigning one of 14 propaganda technique labels to each span). It contains over 500 articles, annotated by experts. The annotations are fine-grained, marking exact character offsets, which allows for a detailed analysis of propaganda within the text.[1]

WANLP [4]   The dataset was developed for the WANLP 2022 Shared Task 3 on Shared Task 3: Propaganda Detection in Arabic Social Media Text. It precisely consists of Arabic news articles annotated for propaganda techniques at the span level. Inspired by the PTC dataset, this is the first of its kind in Arabic, aiming to support fine-grained propaganda detection in a low-resource language.[2]

SemEval 2023 Task 3 [5]   Similarly to [3], this dataset is a benchmark for detecting persuasion techniques in news articles. The main novelty introduced in this version is that it includes news articles in **six languages**, i.e., English, French, German, Italian, Polish, and Russian. The resource comprised data collected from 2020 to mid-2022. It covers a wide range of topics, ranging from COVID-19 to the Russo-Ukrainian war. he dataset draws from both mainstream and alternative media, with many alternative sources flagged by fact-checkers as potential disinformation spreaders. Articles were gathered using news aggregators (e.g., Google News, EMM) and credibility-rating platforms (e.g., NewsGuard, MediaBiasFactCheck).[3]

---

[1] https://huggingface.co/datasets/SemEvalWorkshop/sem_eval_2020_task_11
[2] https://sites.google.com/view/wanlp2022/shared-tasks?authuser=0
[3] https://propaganda.math.unipd.it/semeval2023task3/

BanMANI [6]   This is a novel dataset of 800 Bangla social-media posts paired with 500 reference news articles, annotated for binary manipulation and manipulation spans; 530 instances being manipulated and 270 non-manipulated. The authors propose a semi-automatic annotation pipeline tailored for low-resource NLP settings, ensuring balanced coverage of manipulated versus benign posts.[4]

Salman et al. [7]   This dataset consists of 1,030 English–Roman Urdu code-switched social media snippets, each annotated at the fragment level with one or more of 20 propaganda techniques. This dataset was developed to support a novel task focused on the fine-grained detection of propaganda in multilingual and informal online discourse.[5]

ArAIEval [8]   The ArAIEval Shared Task dataset covers both unimodal (Task 1) and multimodal (Task 2) Arabic content. We focus here on Task 1 data in particular, as it is designed precisely for span-level propaganda detection. The resource comprises 9,000 Arabic text snippets—1,500 tweets and 7,500 news paragraphs—annotated with 23 persuasion techniques, marking precise character-level spans. Annotations were performed by three annotators with expert consolidation across iterative stages.[6]

ArMPro [9]   The ArMPro dataset consists of 20,487 annotated spans across train (15,437), development (1,699), and test (3,351) sets. Each span is labeled with one or more of 23 propaganda techniques.[7]

ZenPropaganda [10]   ZenPropaganda is a Russian-language dataset focused on COVID-19-related online media, containing 125 texts and nearly 2,400 annotated propaganda fragments, labeled with 36 different propaganda techniques. The annotations follow a fine-grained schema inspired by the one used in SemEval 2020, assigning each span a specific propaganda technique.[8]

ManiTweet [11]   The paper focuses on detecting tweets that intentionally misrepresent information from associated news articles. ManiTweet comprises 3,600 tweet-article pairs annotated with binary labels, manipulation spans, and types of manipulation.

4. Discussion

Propaganda detection remains a complex and evolving task, with challenges arising across datasets, models, and evaluation frameworks. From the analysis of these texts, several key factors emerge that are consistently observed across different studies.

First of all, a consistent issue across nearly all corpora is **class imbalance** [12, 10, 9]. Techniques such as `Loaded Language` and `Name Calling` are overrepresented, while rarer strategies (e.g., `Straw Man`, `Whataboutism`) often suffer from near-zero F1-scores [13]. This imbalance impacts model reliability, particularly for zero-shot and few-shot settings with LLMs [14].

Another critical factor that emerges is the variability in **annotation schemes** across datasets. Although span-level annotation, which aims to identify specific persuasion techniques, is the most widely adopted strategy, the degree of granularity and the definition of labels vary considerably. Many datasets build upon the fine-grained labeling introduced by SemEval 2020 Task 11, yet notable divergences remain. For example, ZenPropaganda defines 36 distinct techniques organized into broader meta-classes [10], whereas ArPro adopts a set of 23 techniques and introduces a dedicated `no technique` label for neutral spans [9]. Annotation quality also presents challenges: for example, the ManiTweet dataset [11] requires precise span-level alignment between tweets and their referenced articles, a process inherently affected by subjectivity and interpretative variation.

---

[4]https://github.com/kamruzzaman15/banmani
[5]https://github.com/mbzuai-nlp/propaganda-codeswitched-text
[6]https://gitlab.com/araieval/araieval_arabicnlp24/-/tree/main/task1/data
[7]https://github.com/MaramHasanain/ArMPro
[8]https://github.com/aschern/ru_zen_prop

Table 2
Summary of the datasets' size and type of data instances included for annotation, along with label distribution (when available).

Dataset	Data instances	# instances	Label distribution
PTC [2]	articles	451	18
PTC-SemEval 2020 [3]	articles	536	14
WANLP [4]	tweets	2,000	21
SemEval 2023 Task 3 [5]	articles	2,049	19
BanMANI [6]	articles	800	2
Salman et al. [7]	text snippets	1,030	20
ArAIEval [8]	text snippets, tweets and news paragraphs	9k, 1.5k, 7.5k	23
ArMPro [9]	news articles	2,810	22
ZenPropaganda [10]	text snippets	125	12
ManiTweet [11]	tweets associated to articles	3,636	2

Linguistic diversity is another critical factor. While many datasets remain English-centric, efforts like WANLP [15], ZenPropaganda [10], and ArPro [9] expand the scope to Russian, Arabic, Polish, Bangla, and code-switched content. These studies highlight the limitations of LLMs such as GPT-3.5 and GPT-4 in zero-shot or cross-lingual settings. In particular, BanMANI revealed performance gaps caused by linguistic and cultural mismatches, while ArPro and ArAIEval [14] confirmed that fine-tuned, language-specific models still outperform LLMs on span-based and multi-label tasks.

Another important aspect to highlight is the issue of **domain coverage**. News media remains the most frequently targeted domain, especially in benchmark datasets such as SemEval 2020 [16, 12], SemEval 2023 [17], and ArPro [9]. These resources typically consist of formal texts with span-level annotations of persuasive techniques. However, they vary significantly in geographical and linguistic scope: ArPro focuses on Arabic news, while SemEval 2023 includes six languages, although practical experiments are limited to English and Russian due to tool availability. ZenPropaganda [10], in contrast, narrows the scope to Russian COVID-related media, offering a domain-specific perspective but with limited generalizability.

In contrast, social media presents a markedly different and more challenging environment, characterized by short, informal, and noisy content. Datasets like BanMANI [6] and the code-switched corpus by Salman et al. [7] focus on low-resource and linguistically diverse contexts such as Bangla and English–Roman Urdu, respectively. Similarly, ManiTweet [11] addresses the manipulation of news content within tweets, combining brevity with factual misrepresentation. These examples illustrate the challenges of transferring models trained on formal, monolingual text to informal, multilingual, and often culturally nuanced settings.

When framed within the DCAI paradigm, it is clear that none of the identified datasets provide evidence of life-cycle maintenance or regular updates beyond their initial publication. The static nature of these corpora poses a significant challenge because, as propaganda techniques, languages, and media landscapes evolve, models trained on outdated or unrefreshed data are at risk of poor generalization, reduced robustness, and susceptibility to emerging manipulation strategies. However, the SemEval series and PTC are community benchmarks that have driven progress. Nevertheless, their underlying data has not been periodically re-annotated or augmented to reflect current events or techniques beyond the original release window.

To complement this brief overview, we also examined a selection of research works—identified using the same keywords mentioned in Section 2—that employed these benchmarks to develop and evaluate their own models. The goal is to illustrate the types of models and approaches these datasets have supported, while also highlighting the performance achieved and the main limitations reported in the literature. Table 3 provides a final synthesis of the analyzed works and the corresponding employed datasets for the span identification task, while Table 2 thus summarizes all the information about the datasets identified, models applied, methods employed, limitations encountered, and results achieved in

each cited study.

Table 3
Datasets and corresponding works.

Dataset	Works
SemEval 2020	[12], [16]
SemEval 2023	[12]
BanMANI	[6]
Multiple social networks propaganda 2024	[7]
ZenPropaganda	[10]
WANLP	[15]
ArPro	[9]
Arabic tweets/news 2024	[14]
ManiTweet	[11]

Table 4
Summary of works: methods, models, results, and limitations.

Work (year)	Methods	Models	Results	Limitations
[16]	Chain of thought, few-shot learning	GPT-3.5 and GPT-4	F1 = 0.19	Class imbalance. Irreproducibility. Hallucinations in output.
[12]	Data augmentation. Sentence-Level Feature Concatenation. Oversampling.	BERT	F1 = 0.44. Precision = 0.43. Recall = 0.45	Class imbalance. Computationally expensive.
[12]	Rhetorical Structure Theory	XLM-RoBERTa	micro-F1 = 0.31. macro-F1 = 0.21	Class imbalance. Low scalability. RST complexity.
[6]	Zero-shot experiment	GPT-3 fine-tuned	EM = 12.3	Limited sample size.
[7]	Out-of-Domain, Translated, Code-Switched fine-tuning strategies	Various BERT models, GPT-3.5-Turbo	Micro-F1 = 0.53. Macro-F1 = 0.28. Acc = 0.91	Class imbalance. Limited sample size.
[10]	Adapting and evaluating existing state-of-the-art models	RoBERTa-CRF	F1 = 0.43	Class imbalance. Topic-specific data limitation.
[15]	Data augmentation. Transfer learning	Marefa-NER, XLM-RoBERTa	F1 = 0.78	Class imbalance.
[9]	Fine-tuned PLMs for each task; 5 random seed testing	AraBERT, XLM-RoBERTa, GPT-4	F1 = 0.60 (multi-label), 0.65 (coarse), 0.76 (binary)	Class imbalance.
[14]	Multimodal data fusion	BERT models and visual encoders	Micro-F1 = 0.30	Class imbalance. Multimodality complexity.
[11]	LongFormer-Encoder-Decoder	Various LLMs, LED-FT	F1 = 0.32 (subtask 2), 0.16 (subtask 3)	Dataset not multilingual.

5. Conclusions and Future Work

This mini-review has provided a comprehensive overview of recent research efforts in propaganda detection, with a particular focus on span-level annotation schemes. The analysis of 10 key studies reveals several recurring challenges and patterns. Chief among these is the pervasive issue of class imbalance, which consistently hinders the detection of less frequent persuasion techniques and impacts model robustness, particularly in zero-shot and cross-lingual scenarios. Additionally, the lack of standardization in annotation granularity and label definitions complicates cross-dataset comparisons and model generalization.

Strategies for evolving propaganda detection datasets should emphasize regular updates and continuous improvement. This can be achieved by periodically re-annotating data to capture emerging linguistic trends and evolving propaganda tactics; validating and refining annotations for quality and consistency; and encouraging community-driven contributions to enrich the dataset. Adopting modular dataset designs and leveraging automated monitoring for distributional changes enables flexible adaptation to new challenges. These approaches are essential to ensuring that datasets remain relevant and robust by aligning them with the dynamic nature of language and online propaganda through data-centric AI principles.

In the future, it would be worthwhile to explore several directions. Firstly, it is imperative to develop richer, multilingual, and cross-domain datasets with consistent annotation guidelines to advance model generalization. Greater emphasis should be placed on annotator agreement and validation procedures to ensure reliability and reduce label noise. Secondly, enhancing the adaptability of LLMs through lightweight fine-tuning or hybrid architectures could help narrow the performance disparity with task-specific models, particularly in settings with limited resources. Finally, the incorporation of multimodal data, such as text and images in memes, and context-aware features, such as discourse structures or user metadata, has the potential to further enhance the accuracy of detection in real-world applications, including content moderation, media literacy tools, and automated fact-checking systems.

Acknowledgments

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT to paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Malerba, V. Pasquadibisceglie, Data-centric ai, Journal of Intelligent Information Systems 62 (2024) 1493–1502.

[2] G. Da San Martino, S. Yu, A. Barrón-Cedeño, R. Petrov, P. Nakov, Fine-grained analysis of propaganda in news articles, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 5636–5646. URL: https://aclanthology.org/D19-1565/. doi:10.18653/v1/D19-1565.

[3] G. Da San Martino, A. Barrón-Cedeño, H. Wachsmuth, R. Petrov, P. Nakov, SemEval-2020 task 11: Detection of propaganda techniques in news articles, in: A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, E. Shutova (Eds.), Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics, Barcelona (online), 2020, pp. 1377–1414. URL: https://aclanthology.org/2020.semeval-1.186/. doi:10.18653/v1/2020.semeval-1.186.

[4] F. Alam, H. Mubarak, W. Zaghouani, P. Nakov, G. Da San Martino, Overview of the WANLP 2022 shared task on propaganda detection in Arabic, in: Proceedings of the Seventh Arabic Natural Language Processing Workshop, Association for Computational Linguistics, Abu Dhabi, UAE, 2022.

[5] J. Piskorski, N. Stefanovitch, G. Da San Martino, P. Nakov, SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup, in: A. K. Ojha, A. S. Doğruöz, G. Da San Martino, H. Tayyar Madabushi, R. Kumar, E. Sartori (Eds.), Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 2343–2361. URL: https://aclanthology.org/2023.semeval-1.317/. doi:10.18653/v1/2023.semeval-1.317.

[6] M. Kamruzzaman, M. M. I. Shovon, G. Kim, BanMANI: A dataset to identify manipulated social media news in Bangla, in: A. H. Haddad, A. R. Terryn, R. Mitkov, R. Rapp, P. Zweigenbaum,

S. Sharoff (Eds.), Proceedings of the Workshop on Computational Terminology in NLP and Translation Studies (ConTeNTS) Incorporating the 16th Workshop on Building and Using Comparable Corpora (BUCC), INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 2023, pp. 51–58. URL: https://aclanthology.org/2023.contents-1.7/.

[7] M. U. Salman, A. Hanif, S. Shehata, P. Nakov, Detecting propaganda techniques in code-switched social media text, in: H. Bouamor, J. Pino, K. Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 16794–16812. URL: https://aclanthology.org/2023.emnlp-main.1044/. doi:10.18653/v1/2023.emnlp-main.1044.

[8] M. Hasanain, M. A. Hasan, F. Ahmed, R. Suwaileh, M. R. Biswas, W. Zaghouani, F. Alam, ArAIEval Shared Task: Propagandistic techniques detection in unimodal and multimodal arabic content, in: Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024), Association for Computational Linguistics, Bangkok, 2024.

[9] M. Hasanain, F. Ahmad, F. Alam, Can GPT-4 identify propaganda? annotation and detection of propaganda spans in news articles, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 2724–2744. URL: https://aclanthology.org/2024.lrec-main.244/.

[10] A. Chernyavskiy, S. Shomova, I. Dushakova, I. Kiriya, D. Ilvovsky, ZenPropaganda: A comprehensive study on identifying propaganda techniques in Russian coronavirus-related media, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 17795–17807. URL: https://aclanthology.org/2024.lrec-main.1548/.

[11] K.-H. Huang, H. P. Chan, K. McKeown, H. Ji, ManiTweet: A new benchmark for identifying manipulation of news on social media, in: O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, S. Schockaert (Eds.), Proceedings of the 31st International Conference on Computational Linguistics, Association for Computational Linguistics, Abu Dhabi, UAE, 2025, pp. 11161–11180. URL: https://aclanthology.org/2025.coling-main.739/.

[12] W. Li, S. Li, C. Liu, L. Lu, Z. Shi, S. Wen, Span identification and technique classification of propaganda in news articles, Complex & Intelligent Systems 8 (2022) 3603–3612.

[13] P. N. Ahmad, L. Yuanchao, K. Aurangzeb, M. S. Anwar, Q. M. u. Haq, Semantic web-based propaganda text detection from social media using meta-learning, Service Oriented Computing and Applications (2024) 1–15.

[14] M. Hasanain, M. A. Hasan, F. Ahmad, R. Suwaileh, M. R. Biswas, W. Zaghouani, F. Alam, ArAIEval shared task: Propagandistic techniques detection in unimodal and multimodal Arabic content, in: N. Habash, H. Bouamor, R. Eskander, N. Tomeh, I. Abu Farha, A. Abdelali, S. Touileb, I. Hamed, Y. Onaizan, B. Alhafni, W. Antoun, S. Khalifa, H. Haddad, I. Zitouni, B. AlKhamissi, R. Almatham, K. Mrini (Eds.), Proceedings of the Second Arabic Natural Language Processing Conference, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 456–466. URL: https://aclanthology.org/2024.arabicnlp-1.44/. doi:10.18653/v1/2024.arabicnlp-1.44.

[15] A. S. Hussein, A. B. S. Mohammad, M. Ibrahim, L. H. Afify, S. R. El-Beltagy, NGU CNLP at WANLP 2022 shared task: Propaganda detection in Arabic, in: H. Bouamor, H. Al-Khalifa, K. Darwish, O. Rambow, F. Bougares, A. Abdelali, N. Tomeh, S. Khalifa, W. Zaghouani (Eds.), Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP), Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 2022, pp. 545–550. URL: https://aclanthology.org/2022.wanlp-1.66/. doi:10.18653/v1/2022.wanlp-1.66.

[16] J. Szwoch, M. Staszkow, R. Rzepka, K. Araki, Limitations of large language models in propaganda detection task, Applied Sciences 14 (2024) 4330.

[17] A. Chernyavskiy, D. Ilvovsky, P. Nakov, 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), 2024, pp. 1452–1462.