

Legend at ArAIEval Shared Task: Persuasion Technique Detection using a Language-Agnostic Text Representation Model

Olumide E. Ojo^{1,5,a}, Olaronke O. Adebajji^{1,b}, Hiram Calvo^{1,c}, Damian O. Dieke^{3,d},
Olumuyiwa E. Ojo^{4,e}, Seye E. Akinsanya^{2,f}, Tolulope O. Abiola^{2,g}, Anna Feldman^{5,h}

¹Instituto Politécnico Nacional, Centro de Investigación en Computación, CDMX, Mexico

²Federal University Oye-Ekiti, Nigeria; ³Caritas University, Nigeria; ⁴Lead City University, Nigeria;

⁵Montclair State University, USA

{^aolumideoea, ^bolaronke.oluwayemisi, ^chiramcalvo, ^ddieketobesky, ^emuyeskin, ^fakinsanyaseye, ^gabiolato92}@gmail.com;

^hfeldmana@montclair.edu

Abstract

In this paper, we share our best performing submission to the Arabic AI Tasks Evaluation Challenge (ArAIEval) at ArabicNLP 2023. Our focus was on Task 1, which involves identifying persuasion techniques in excerpts from tweets and news articles. The persuasion technique in Arabic texts was detected using a training loop with XLM-RoBERTa, a language-agnostic text representation model. This approach proved to be potent, leveraging fine-tuning of a multilingual language model. In our evaluation of the test set, we achieved a micro F1 score of 0.64 for subtask A of the competition.

1 Introduction

In an era defined by the rapid dissemination of information through digital channels, the task of recognizing persuasion techniques in text is now more crucial than ever (Hasanain et al., 2023; Hosain et al., 2021; Gupta et al., 2021; Sadeghi et al., 2023; Alam et al., 2022; Abujaber et al., 2021). The advent of the Internet and social media has created new avenues for influence and manipulation (Dholakia et al., 2023; Ruffo et al., 2023; Botes, 2023). Although these technological advances have undoubtedly given people unparalleled access to information and a platform to express their thoughts, they have also introduced new avenues for persuasion, influence, and even propaganda. At the heart of this shift in our way of life is the fundamental challenge of distinguishing between genuinely informative and impartial content and content subtly crafted to promote a specific agenda or ideology. A critical component of the fight against spreading misinformation is the development of tools and resources in NLP that can detect persuasion technique in news articles and posts on social media.

Arabic language is among the most spoken languages in the world (Ghazzawi, 1992). The Arabic speaking world stands out in importance due to

its intricate mix of language, culture, and geography. The diversity of linguistic expressions in Arabic extends beyond the spoken word and permeates every aspect of life, including digital text. Arabic connects a vast and diverse population of native speakers and foreigners with its rich heritage and multiple dialects. Its influence extends over a wide territory, from Arab nations in the southern part of the Arabian Peninsula, to Asia, and to the Maghreb in North Africa and the heart of the Arabian Peninsula (Huafeng et al., 2019). In this digital age, the importance and extensive use of Arabic in various forms reflects the profound impact of technology on this linguistic community. It has provided a platform for Arabic speakers around the world to engage in dialogue, share ideas, and express their thoughts in a global context. The diverse Arab-speaking populations foster a rich and dynamic environment where individuals can connect, collaborate, and debate issues of global significance. However, the proliferation of digital media in Arabic also presents challenges. Digital media have become a fertile ground for the dissemination of persuasive content, including propaganda, misinformation, and various forms of manipulation (Aleroud et al., 2023; Abd Elaziz et al., 2023). The use of technology in preserving the integrity of the Arabic language and ensuring responsible use of digital media is of utmost importance.

In today's world, we are surrounded by information, especially on the Internet and social media. Text classification can serve as a foundational step for the detection of the persuasion technique in text on social media and the Internet. These texts can be classified according to their emotional tone using sentiment analysis techniques (Nikolaidis et al., 2023; Ojo et al., 2022b, 2021, 2020, 2023, 2022a; Piskorski et al., 2023; Hromadka et al., 2023). Persuasion techniques can be tricky to spot because they come in many forms, such as stories, logical arguments, or even subtle language tricks to

sway our thinking. These techniques are powerful and are not always used for good reasons. To deal with this, researchers and experts are working on ways to detect when someone is trying to persuade people through text. In this way, we will be able to recognize when we are influenced and when people are spreading false or misleading information, which can be harmful. In this paper, we develop a methodology to automatically identify and analyze persuasion techniques in text using the XLM-RoBERTa model. Using the power of NLP, we can harness the capabilities of state-of-the-art models and unravel the persuasion techniques embedded within Arabic text, contributing to both media ethics and the combating of misinformation. The intricacies of persuasion technique detection in Arabic are discussed in detail, along with the possibility of applying this knowledge across different languages.

2 Related Work

A significant amount of research has been conducted on the detection of persuasion techniques in text (Hasanain et al., 2023; Modzelewski et al., 2023; Hossain et al., 2021; Sadeghi et al., 2023; Abujaber et al., 2021). Researchers have explored binary and multilabel approaches to detecting persuasion techniques.

In their article, the authors in (Modzelewski et al., 2023) focused on detecting genres and persuasion techniques in multiple languages using various data augmentation techniques to enhance their models. For genre detection, they created synthetic texts using the GPT-3 Davinci language model, while for persuasion technique detection, they augmented the dataset using text translation with the DeepL translator. Their fine-tuned models achieved top ten rankings in all languages, demonstrating the effectiveness of their approach. They also excelled in genre detection, securing top positions in Spanish, German, and Italian. They presented a single multilingual system using the RoBERTa model to classify online news genres and improved this system by adding texts generated by the GPT3-Davinci model to the training dataset.

(Maram et al., 2023) addressed misinformation in mainstream and social media and the challenges faced by manual detection and verification efforts by journalists and fact checkers in the SemEval-2023 task (Piskorski et al., 2023). The task included three subtasks, six languages, and three

surprise test languages, totaling 27 different test scenarios. The authors successfully submitted entries for all 27 test setups and the official results placed their system among the top three for 10 of these setups. They fine-tuned transformer models in the multiclass and multilabel classification settings, experimenting with both monolingual and multilingual pre-trained models, as well as data augmentation. Their multilingual model based on XLM-RoBERTa demonstrated superior performance in all tasks, even for languages not seen during training.

The most effective solution in the detection of persuasion techniques for Subtask 3 of SemEval 2023 Task 3 by (Hromadka et al., 2023) delivered promising performance, with micro-F1 scores ranging from 36 to 55% for languages seen during training and 26 to 45% for languages unseen. Given the multilingual nature of the data and the presence of 23 labels (resulting in limited labeled data for some language-label combinations), the authors chose to fine-tune pre-trained transformer-based language models. Through extensive experimentation, they identified the optimal configuration, featuring a large multilingual model (XLM-RoBERTa) trained on all input data, with carefully calibrated confidence thresholds for known and surprise languages separately. Their final system demonstrated superior performance, ranking first in six of nine languages, including two surprise languages, and achieving highly competitive results in the remaining three languages.

The experiments of (Nikolaidis et al., 2023) focused on the detection of persuasion techniques in online news articles in Polish and Russian. These experiments used a taxonomy comprising 23 distinct persuasion techniques. Persuasion techniques were evaluated in several ways, including the granularity of the classification (coarse with six labels or fine with 23 labels) and the level of location of the labels (subword, sentence, paragraph). The study compared the performance of state-of-the-art transformer-based models trained both monolingually and multilingually. The findings indicate that multilingual models generally outperform monolingual models in various evaluation scenarios. However, due to the complexity of the task, there remains substantial room for improvement in the field of persuasion technique detection within online news articles.

Inspired by (Maram et al., 2023; Hromadka et al.,

2023), our research focus is to determine whether a multi-genre snippet (tweets and news paragraphs of news articles) contains a persuasion technique or not. This is a binary classification task, and we are categorizing Arabic text either as containing a persuasion technique or as lacking persuasion technique. Our approach involves fine-tuning a large pre-trained language model based on transformers. We conducted experiments with different language models and concluded with XLM-RoBERTa due to its superior performance. Furthermore, we fine-tuned our system by tweaking hyperparameters and conducting multiple iterations. This process involved calculating cross-entropy loss, performing backpropagation, and updating the model's weights.

3 Persuasion Technique Detection

3.1 Dataset Analysis

Datasets provided by the organizers consist of a diverse collection of Arabic text samples, each associated with a label. These texts have been carefully selected to represent where persuasion techniques are present or not. The dataset were labeled as 'true' or 'false', where the true class represents texts where persuasion techniques are used to influence the reader's opinion, and the false class consists of texts that do not use persuasion techniques. The method involves binary classification to determine the presence of a persuasion technique within the document.

The dataset comprises 2,427 samples in the training set and 503 in the test set. Within the training data, there is an imbalance in the label distribution, with 1,918 samples labeled as "true" and 509 labeled as "false". An example of text that represents the persuasion technique in the dataset is shown in Figure 1.

```
id: 00425
text: حريق ميبيت معهد تالة: إهمال، شبهة فساد وأزمة ثقة [فيديو]
label: true
type: tweet
```

Figure 1: Persuasion technique in the Arabic text

3.2 Application of XLM-RoBERTa

Leveraging the Hugging Face Transformers library, we prepared and formatted the data so that it could be seamlessly incorporated into our model. In line

with the methodology outlined in Section 4, we designed a custom model architecture and then proceeded to encode and load the data for further processing. During training, we optimized the model's classification head to minimize cross-entropy loss, and predictions are made on unseen text, followed by post-processing. This approach demonstrates the adaptability and effectiveness of XLM-RoBERTa in diverse linguistic contexts, offering practical solutions to automate the detection of persuasion techniques across languages.

4 System Setup and Experiments

4.1 Training Strategy

Our training strategy involves fine-tuning the XLM-RoBERTa model using the provided dataset. To optimize the performance of the model, we incorporate several key components and hyperparameters, which are summarized in Table 1.

4.2 Model Fine-Tuning

During model fine-tuning, we added a classification layer at the end of the pre-trained XLM-RoBERTa model. This additional layer allows the model to perform the specific classification task required. To prevent overfitting, a dropout layer was incorporated.

4.3 Class Weights

To address the class imbalance in the dataset, we adjusted the learning process using class weights. This ensures that the model effectively learns from all classes and is not biased by data imbalance.

4.4 Evaluation Metric

The model's performance was evaluated using the micro-score F1, the default metric for this subtask. This metric provides a comprehensive measure of the model's classification performance.

4.5 Training Process

The training process involved conducting a total of 6 epochs, where each epoch represents a complete pass through the entire training dataset. To avoid overfitting, early stopping was employed on the basis of cross-entropy loss.

4.6 Learning Rate Scheduler

A learning rate scheduler was implemented to dynamically adjust the learning rate during training.

Hyperparameter	Value
Learning rate	5×10^{-5}
Batch size	16
Epochs	6
Optimizer	ADAM
Early stopping	Cross-entropy loss
Learning rate scheduler	StepLR (factor=0.85, step size=2)

Table 1: Model Hyperparameters

Specifically, we used a StepLR scheduler with a reduction factor of 0.85 applied every 2 epochs. This scheduling strategy contributes to training stability and controlled convergence.

4.7 Optimization Process

To initiate the optimization process, we reset the gradients (setting them to zero). Subsequently, we conducted backpropagation to calculate gradients for all model parameters and, lastly, update the model’s parameters using the computed gradients.

Our approach leverages these components and hyperparameters to fine-tune the model effectively, ensuring robust and controlled training.

5 Results

The evolution of pre-trained language models has ushered in significant advancements across different NLP tasks. In this section, we present the results of our experiments on persuasion technique detection in Arabic text. Our findings revealed the versatility of XLM-RoBERTa in effectively handling multilingual data. We evaluated the model using the micro-F1 score and an overview of the results achieved by our model on the dataset is shown in Table 2.

Model	F1-Score
Baseline model	0.4771
XLM-RoBERTa	0.6402

Table 2: Performance Comparison of Models for Persuasion Technique Detection in Arabic Text

From Table 1, our findings reveal the effectiveness of XLM-RoBERTa in identifying persuasion techniques within the Arabic text. The model’s pre-trained knowledge, combined with its cross-lingual capabilities, makes it a promising tool for similar tasks in other languages as well. As Arabic

is a morphologically rich language with various dialects, the success of XLM-RoBERTa in this task is a testament to its robustness and versatility.

6 Conclusion

The application of XLM-RoBERTa, a language-agnostic text representation model, for the detection of the persuasion technique in Arabic text illustrates the growing potential of cross-lingual models in specialized NLP tasks. Arabic, with its rich morphology and diverse dialects, presents unique challenges for text analysis. Our proposed model has the ability to capture the underlying structure and semantics of persuasion technique in text, regardless of language. The results obtained in our analysis demonstrate that XLM-RoBERTa can adapt effectively and perform well on such intricate tasks, even in languages that are structurally different from the ones they were originally trained on. This not only underscores the versatility of XLM-RoBERTa but also sets a promising direction for further research in detecting persuasion techniques across various languages. In future work, we plan to accommodate more languages in the dataset, and fine-tune other multilingual models for this task.

Ethics Statement

We acknowledge the influence that scientific research can have on society and recognize our responsibility to ensure its positive impact. Our research is carried out with the aim of addressing real-world challenges, promoting well-being, and improving quality of life. We actively seek feedback and input from those affected by our research. In the event that our research may have unintended negative impacts, we are committed to addressing and rectifying those issues promptly.

Acknowledgments

This work was done with partial support from the Mexican Government through the grant A1-S-47854 of CONACYT, Mexico, grants 20232138, 20230140, 20232080 and 20231567 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico and acknowledge the support of Microsoft through the Microsoft Latin America PhD Award.

References

- Mohamed Abd Elaziz, Abdelghani Dahou, Dina Ahmed Orabi, Samah Alshathri, Eman M Soliman, and Ahmed A Ewees. 2023. A hybrid multitask learning framework with a fire hawk optimizer for arabic fake news detection. *Mathematics*, 11(2):258.
- Dia Abujaber, Ahmed Qarqaz, and Malak A Abdullah. 2021. Lecun at semeval-2021 task 6: detecting persuasion techniques in text using ensemble pretrained transformers and data augmentation. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 1068–1074.
- Firoj Alam, Hamdy Mubarak, Wajdi Zaghouni, Giovanni Da San Martino, and Preslav Nakov. 2022. [Overview of the WANLP 2022 shared task on propaganda detection in Arabic](#). In *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 108–118, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
- Ahmed Aleroud, Abdullah Bani Melhem, Nour Alhussien, and Craig Douglas Albert. 2023. span-prop: Combatting contextualized social media state-linked propaganda in the middle east.
- Marietjie Botes. 2023. Autonomy and the social dilemma of online manipulative behavior. *AI and Ethics*, 3(1):315–323.
- Nikhilesh Dholakia, Aras Ozgun, and Deniz Atik. 2023. The miasma of misinformation: a social analysis of media, markets, and manipulation. *Consumption Markets & Culture*, pages 1–16.
- Sabah Ghazzawi. 1992. The arabic language. *Washington DC: Center for Contemporary Arab Studies*.
- Kshitij Gupta, Devansh Gautam, and Radhika Mamidi. 2021. Volta at semeval-2021 task 6: Towards detecting persuasive texts and images using textual and multimodal ensemble. *arXiv preprint arXiv:2106.00240*.
- Maram Hasanain, Firoj Alam, Hamdy Mubarak, Samir Abdaljalil, Wajdi Zaghouni, Preslav Nakov, Giovanni Da San Martino, and Abed Alhakim Freihat. 2023. ArAIEval Shared Task: Persuasion Techniques and Disinformation Detection in Arabic Text. In *Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023)*, Singapore. Association for Computational Linguistics.
- Tashin Hossain, Jannatun Naim, Fareen Tasneem, Radiathun Tasnia, and Abu Nowshed Chy. 2021. Csecudsg at semeval-2021 task 6: Orchestrating multimodal neural architectures for identifying persuasion techniques in texts and images. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 1088–1095.
- Timo Hromadka, Timotej Smolen, Tomas Remis, Branislav Pecher, and Ivan Srba. 2023. Kinitveraaai at semeval-2023 task 3: Simple yet powerful multilingual fine-tuning for persuasion techniques detection. *arXiv preprint arXiv:2304.11924*.
- Hu Huafeng, Fan Zhihao, and Gao Xuezheng. 2019. The history of the arab nation and the arabic language. *Arabic Language, Literature & Culture*, 4(3):48.
- Hasanain Maram, El-Shangiti Ahmed Oumar, Nandi Rabinindra Nath, Nakov Preslav, and Alam Firoj. 2023. Qcri at semeval-2023 task 3: News genre, framing and persuasion techniques detection using multilingual models. *arXiv preprint arXiv:2305.03336*.
- Arkadiusz Modzelewski, Witold Sosnowski, Magdalena Wilczynska, and Adam Wierzbicki. 2023. Dshacker at semeval-2023 task 3: Genres and persuasion techniques detection with multilingual data augmentation through machine translation and text generation. In *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1582–1591.
- Nikolaos Nikolaidis, Nicolas Stefanovitch, and Jakub Piskorski. 2023. On experiments of detecting persuasion techniques in polish and russian online news: Preliminary study. In *Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023)*, pages 155–164.
- OE Ojo, A Gelbukh, H Calvo, and OO Adebajji. 2021. Performance study of n-grams in the analysis of sentiments. *Journal of the Nigerian Society of Physical Sciences*, pages 477–483.
- OE Ojo, A Gelbukh, H Calvo, A Feldman, OO Adebajji, and J Armenta-Segura. 2022a. Language identification at the word level in code-mixed texts using character sequence and word embedding. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts*, pages 1–6.
- Olumide E Ojo, Alexander Gelbukh, Hiram Calvo, Olaronke O Adebajji, and Grigori Sidorov. 2020.

- Sentiment detection in economics texts. In *Mexican International Conference on Artificial Intelligence*, pages 271–281. Springer.
- Olumide Ebenezer Ojo, Thang Ta Hoang, Alexander Gelbukh, Hiram Calvo, Grigori Sidorov, and Olaronke Oluwayemisi Adebajji. 2022b. Automatic hate speech detection using cnn model and word embedding. *Computación y Sistemas*, 26(2).
- Olumide Ebenezer Ojo, Hoang Thang Ta, Alexander Gelbukh, Hiram Calvo, Olaronke Oluwayemisi Adebajji, and Grigori Sidorov. 2023. Transformer-based approaches to sentiment detection. In *Recent Developments and the New Directions of Research, Foundations, and Applications: Selected Papers of the 8th World Conference on Soft Computing, February 03–05, 2022, Baku, Azerbaijan, Vol. II*, pages 101–110. Springer.
- 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 the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361.
- Giancarlo Ruffo, Alfonso Semeraro, Anastasia Giachanou, and Paolo Rosso. 2023. Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. *Computer science review*, 47:100531.
- Aryan Sadeghi, Reza Alipour, Kamyar Taeb, Parimehr Morassafar, Nima Salemahim, and Ehsaneddin Asgari. 2023. Sinaai at semeval-2023 task 3: A multilingual transformer language model-based approach for the detection of news genre, framing and persuasion techniques. In *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2168–2173.