UnedMediaBiasTeam @ SemEval-2023 Task 3: Can We Detect Persuasive Techniques Transferring Knowledge From Media Bias Detection?

Francisco-Javier Rodrigo-Ginés[1,a], **Laura Plaza**[1,2,b], and **Jorge Carrillo-de-Albornoz**[1,2,b]

[1]NLP & IR Group, UNED / Madrid, 28040, Spain
[2]RMIT University, 3000 Melbourne, Australia
[a]*frodrigo@invi.uned.es*
[b]*{lplaza, jcalbornoz}@lsi.uned.es*

Abstract

How similar is the detection of media bias to the detection of persuasive techniques? We have explored how transferring knowledge from one task to the other may help to improve the performance. This paper presents the systems developed for participating in the SemEval-2023 Task 3: *Detecting the Genre, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup*. We have participated in both the subtask 1: *News Genre Categorisation*, and the subtask 3: *Persuasion Techniques Detection*. Our solutions are based on two-stage fine-tuned multilingual models. We evaluated our approach on the 9 languages provided in the task. Our results show that the use of transfer learning from media bias detection to persuasion techniques detection is beneficial for the subtask of detecting the genre (macro F1-score of 0.523 in the English test set) as it improves previous results, but not for the detection of persuasive techniques (micro F1-score of 0.24 in the English test set).

1 Introduction

Misinformation, and propaganda in particular, represents a major issue in online media. The ability to detect and analyze persuasive techniques in text is essential in order to be able to identify and counter misinformation. This task aims to foster the use of Artificial Intelligence to perform media analysis in order to address this issue.

The task of automated persuasive techniques detection has been an active research area in recent years (Martino et al., 2020b). Previous works have focused on different aspects of the problem, such as the development of datasets (Martino et al., 2020a), the development of persuasive techniques taxonomies (Jia, 2020) and the development of automated systems for the detection of persuasive techniques (Yoosuf and Yang, 2019).

The SemEval-2023 task 3 (Piskorski et al., 2023) focuses on the detection of genre, framing and persuasion techniques in news articles in a multilingual setup. The task consists of three subtasks: News Genre Categorisation, Framing Detection and Persuasion Techniques Detection. The data presented in the task is both multilabel and multilingual, and it also covers complementary dimensions of what makes text persuasive, namely style and framing. The task covers the following languages: English, French, German, Italian, Polish, Russian, Spanish, Greek and Georgian.

We propose to use a two-stage fine-tuning (ValizadehAslani et al., 2022) of multilingual models. We use the XLM-RoBERTa (Conneau et al., 2019) model for subtask 1, and BERT-multilingual (Devlin et al., 2018) model for subtask 3. The reason why these models are used can be seen in the Subsection on ablation analysis. Both models have been fine-tuned on the BABE (Spinde et al., 2022) and MBIC (Spinde et al., 2021) datasets for the task of media bias detection. We then fine-tune the model on the data provided for the task to adapt it to the specific task.

Our approach achieves good results in the first subtask, and average results on the third. We have obtained a macro F1-score of 0.523 (position 8 out of 22) for the English test set in the first subtask, and a micro F1-score of of 0.24 (position 18 out of 23) for the English test set in the third subtask. We performed an error and ablation analysis to analyze the errors of our system, and for analyzing the effect of different techniques on the performance. Our system struggles with the detection of more complex persuasion techniques, especially those less represented in the dataset.

To promote replicability, both the code and the models are publicly available. The code can be found at GitHub[1], and the models can be found at HuggingFace: subtask 1[2], subtask 3[3].

[1]https://link.franrodrigo.es/SemEval23-task3-git
[2]https://link.franrodrigo.es/SemEval23-task3-model1
[3]https://link.franrodrigo.es/SemEval23-task3-model3

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 787–793

2 Background

The task of media bias detection is closely related to the task of persuasive techniques detection. It can be seen as a wider problem that encompasses the detection of various persuasive techniques. As the systematic review of (Rodrigo-Ginés et al., under review) lists, certain forms of media bias coincide with the persuasive techniques that are part of this task: appeal to authority, appeal to groupthink/popularity, red herring, loaded language, and labeling. It is therefore natural to explore the transfer of knowledge from one task to the other in order to improve the performance of automatic detection systems, as by leveraging knowledge from the source task, the model can potentially learn faster, require less labeled data, and achieve better performance on the target task.

The SemEval-2023 task 3 (Piskorski et al., 2023) is organized as a multi-lingual task with 3 subtasks: News Genre Categorization, Framing Detection and Persuasion Techniques Detection. The task focuses on online news articles in 9 languages: English, French, German, Italian, Polish, Russian, Spanish, Greek and Georgian. Our efforts have focused on the first and third tasks.

The News Genre Categorization subtask consists of determining the genre of a given article (opinion, objective news or satire). The task is framed as a multi-class classification task at article-level. The Persuasion Techniques Detection subtask aims to determine the persuasion techniques used in each paragraph of an article. The task is framed as a multi-label classification task at paragraph-level.

The main contribution of this work is the use of transfer learning from media bias detection to persuasion techniques detection. To the best of our knowledge, this is the first time this knowledge transfer has been explored for this task.

In recent years, researchers in media bias detection have explored different approaches to generalize media bias. Two categories of approaches have emerged: non-neural network models and neural network models. Non-neural network models rely on classical machine learning methods, which require some feature engineering to generate linguistic (Hube and Fetahu, 2018) or reported speech (Lazaridou and Krestel, 2016) features. On the other hand, neural network models have been shown to outperform traditional methods. In particular, RNNs (Rashkin et al., 2017) and transformers (Baly et al., 2020) are the most commonly used neural networks for media bias detection.

We have fine-tuned some transformers models, using the MBIC (Spinde et al., 2021) and BABE datasets (Spinde et al., 2022). These datasets are designed specifically for the task of media bias detection and contain annotated articles from various sources in multiple classes.

The use of these datasets allow us to leverage the pre-trained models to capture the language-independent features of media bias and to improve the performance of our systems by taking advantage of the knowledge acquired in other tasks. As these datasets only contain texts written in English, we do not expect them to have a direct impact on the performance of the systems for the other languages, but we expect them to help the model to learn language-independent features.

3 System Overview

3.1 Exploratory Data Analysis

The dataset is divided into three subsets: training, development and test. The training set contains 1,234 articles. The development set is used to evaluate the systems during the training phase and consists of 358 articles. The test set is used for the final evaluation and consists of 547 articles.

The data is highly unbalanced across classes. In the first subtask, the majority of the articles belong to the category of *opinion news*. In the third subtask, the most frequent classes are *loaded language*, *labeling* and *doubt*, as it can be seen in Figure 1.

The imbalance of the classes is the main challenge of the task, as the models need to be able to detect and classify rare classes. To minimize this problem we tried various techniques, including under-sampling the majority classes, oversampling the minority ones, augmenting the data creating new instances, and using the multiple loss functions. The different loss functions we have used in order to train our models are: cross entropy loss function, which is very common in multi-class classification problems; the focal loss, which is a variation of the cross entropy loss function designed to deal with unbalanced data (Qin et al., 2018); and the DICE loss function (Li et al., 2019), which is sometimes more suitable for unbalanced data.

After various experiments, we decided to stick with cross entropy along with over/under-sampling techniques. In order to re-balance the class distributions when sampling from the unbalanced dataset, and estimate the sampling weights automatically,

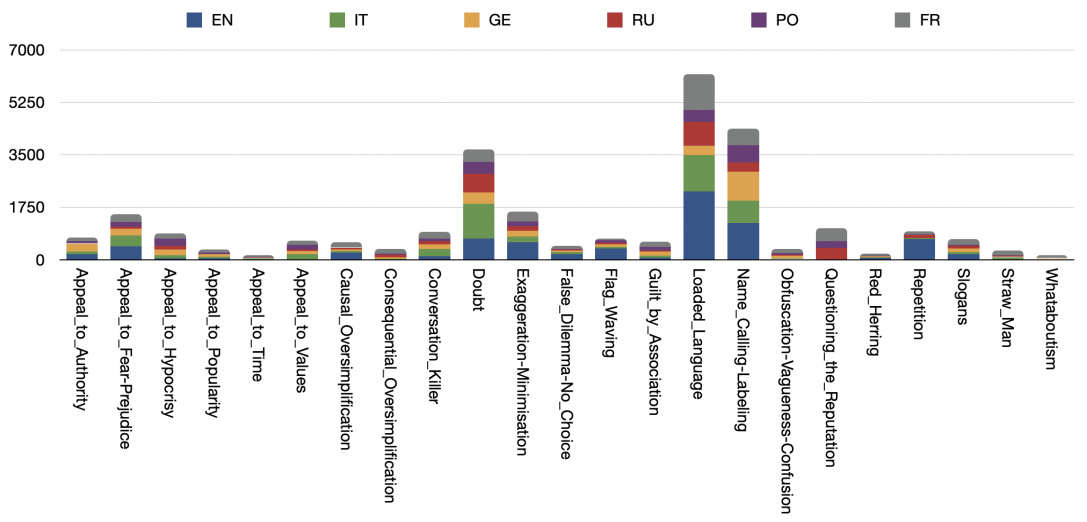

Figure 1: Number of instances per class for subtask 1.

we used the torchsampler library. We discarded the use of data augmentation (using the NLPaug library), since it favors over-training.

3.2 Use of MBIC and BABE Datasets

We decided to leverage the knowledge acquired from the MBIC and BABE datasets (Spinde et al., 2021, 2022) for the task of media bias detection. We used these datasets to fine-tune the models for the task. The MBIC and BABE datasets are designed for the task of media bias detection and contain articles from different sources in English. Both datasets have multiple labels, we have used the label opinion ("Somewhat factual but also opinionated", "Expresses writer's opinion", or "No agreement") for the first subtask, and the label bias ("Biased", "Non-biased") for the second.

3.3 Subtask 1: News Genre Categorization

As it can be seen in Figure 2, our system for the first subtask is based on a two-stage fine-tuning of the XLM-RoBERTa model. We first fine-tune the model on the MBIC and BABE datasets using the cross-entropy loss function. We then fine-tune the model on the data provided for the task to adapt it to the specific task.

By fine-tuning the model on the MBIC and BABE datasets in the first stage, the model can learn to recognize media bias in a more general sense. This allows the model to better identify biased language, rhetoric, and other features that may be present in the data provided for the specific task.

For the fine-tuning on the data provided for the task, we have used a learning rate of 5e-5 and a batch size of 32. We also applied the AdamW optimizer with a weight decay of 0.01. We used a maximum number of epochs of 10, with an early stopping if the model does not improve its F1-score after 3 consecutive epochs.

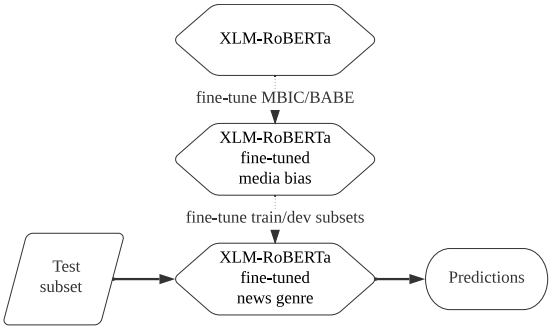

Figure 2: Subtask 1 system overview.

We did several tests, and this fine-tuning setup made in two phases improves the results obtained with systems trained in a single phase.

3.4 Subtask 3: Persuation Techniques Detection

The system developed for multi-label and multi-class detection of persuasion techniques is similar to the one developed for the first task. In this case, instead of training a single model in two phases, we train two models and carry out the cascading inference. The first model is a BERT-multilingual fine-tuned on the MBIC and BABE datasets in a similar way as for the first subtask. The second model is a BERT-multilingual model that has been fine-tuned on the data provided for the task.

Once the two models have been trained, we per-

form the cascading inference: the first model is used to get biased v. non-biased predictions, and the second model is used to get the persuasion predictions. We have used a threshold of 0.35 to decide if an article is biased or not, we arrived at this value after optimizing the threshold in the development phase. If the article is labeled as biased, then the persuasion techniques predictions of the second model are taken into account. If not, then the predictions of the second model are ignored. In Figure 3, we show an overview of the system.

We have also experimented with a different approach, where the predictions of the first model are used as features for the second model. This approach yielded worse results than the cascading approach. We believe this is due to the fact that the models are not able to capture the complex interactions between the two tasks and that the cascading approach is more suitable for this task.

We have used a learning rate of 5e-5 and a batch size of 8 for the fine-tuning of the second model. We have also applied the AdamW optimizer with a weight decay of 0.01. We have used a maximum number of epochs of 5.

Table 1: Subtask 3 classification report (on dev subset) for English

Class	F1-score	Support
Appeal to Authority	0.32	170
Appeal to Fear	0.34	379
Appeal to Hypocrisy	0.20	221
Appeal to Popularity	0.00	110
Appeal to Time	0.00	42
Appeal to Values	0.32	193
Causal Oversimplif.	0.14	111
Conseq. Oversimplif.	0.18	95
Conversation Killer	0.27	241
Doubt	0.53	865
Exaggeration/Minim.	0.28	347
False Dilemma	0.04	132
Flag Waving	0.43	174
Guilt by Association	0.31	115
Loaded Language	0.65	1349
Name Calling/Labeling	0.63	941
Vagueness/Confusion	0.00	96
Quest. the reputation	0.43	440
Red Herring	0.00	44
Repetition	0.19	211
Slogans	0.28	132
Straw Man	0.10	61
Whataboutism	0.00	35
micro average	0.45	6504
macro average	0.25	6504

4 Results and Error Analysis

4.1 Experimental Setup

The data provided for the task consists of training and development sets for the 9 languages provided. The development set is used to evaluate the systems and is not annotated. The task also provides an online submission website to evaluate the systems.

For the training and development sets, we use a stratified split, where the data is split into training and development sets in a stratified manner, with the same proportions of classes in both sets.

We used the Huggingface's transformers library for pre-processing and training of the models. The code was written using Python 3.8 and PyTorch 1.7.1. The models were trained on two NVIDIA GeForce RTX 2080 Ti GPUs.

Finally, we used the proposed metrics for each subtask: macro F1-score for the subtask of News Genre Categorisation, and the micro F1-score for the subtask of Persuasion Techniques Detection.

We evaluated our system on the 9 languages provided in the task, but as we trained two systems based on fine-tuning on datasets with instances only in English, our analysis of results is based on the results obtained for this language. At any rate, we have obtained similar results in all the languages.

We achieved good results in the first task, and average results in the second. In the subtask of News Genre Categorisation, our systems obtained a macro F1-score of 0.523 in the English test set (eighth position, the best system obtained a score of 0.784). In the subtask of Persuasion Techniques Detection, our systems obtained a micro F1-score of 0.24 in the English test set (eighteenth position, the best system obtained a score of 0.375). The detailed results by language and task are listed in Appendix A.

4.2 Ablation Analysis

We have studied various models based on transformers for each of the subtasks for the English subset. The models evaluated were: Multilingual-MiniLM-L12, BERT-base-multilingual, and XLM-RoBERTa-base-multilingual. We also evaluate the impact of the transfer learning from the MBIC and BABE datasets. For the first subtask we found that the best performing model is the XLM-RoBERTa-base-multilingual model. This model was also the best performing model when fine-tuned on the MBIC and BABE datasets. See Table 2.

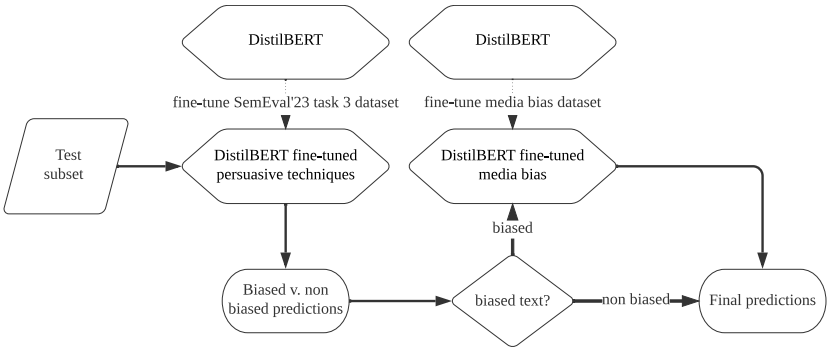

Figure 3: Subtask 3 system overview.

Table 2: Ablation analysis

Model for subtask 1	**F1-score**	**F1-score w. MBIC/BABE**
XLM-RoBERTa-multiling.	0.547	0.624
BERT-multilingual	0.513	0.591
Multilingual-MiniLM-L12	0.492	0.556
Model for subtask 3	**F1-score**	**F1-score w. MBIC/BABE**
XLM-RoBERTa-multiling.	0.231	0.237
BERT-multilingual	0.228	0.243
Multilingual-MiniLM-L12	0.173	0.189

For the subtask of Persuasion Techniques Detection, we found that the best performing model is the XLM-RoBERTa-base-multilingual model. However, the best performing model when fine-tuned on the MBIC and BABE datasets is again the BERT-base model.

Once the base models were chosen, we tested several loss functions for the subtask 1, obtaining the following results: a macro F1-score of 0.624 for the cross-entropy loss function, a macro F1-score of 0.592 for the local loss function, and a macro F1-score of 0.601 for the DICE loss function. We maintained the cross-entropy loss function for the subtask 3 as it was the most suitable for the multi-label problem. Finally, we created some instances based on data augmentation by synonym replacement, but the results worsened (F1=0.587 with data augmentation v. F1=0.624 without it).

4.3 Error analysis

We performed an error analysis to identify the types of errors made by our system. In order to do so, we looked for articles that were misclassified by our system looking at the confusion matrix, and inspecting random samples of articles that were misclassified. We found that our system struggles with the classes less represented in the dataset for both subtasks.

We found some errors related to the task of News Genre Categorisation. Our system often confuses *satire* texts with *opinion news* texts. This is due to the fact that some of the satirical texts in the dataset contain opinionated claims, which makes them difficult to differentiate from opinion news.

In the subtask of Persuasion Techniques Detection, as it can be seen in Table 1, our system struggles with the detection of more persuasion techniques with less support, that is, techniques with less examples in the dataset. This is especially true for the persuasion techniques that are less represented in the dataset, such as *vagueness*, *red herring*, *straw man* and *whataboutism*. We also believe that this may happen due to the fact that the models have difficulty in learning the subtle differences between these techniques.

5 Future work and conclusions

The results obtained shows that the use of transfer learning from media bias detection to persuasive techniques detection is beneficial for the subtask of detecting the genre. In the case of the detection of the persuasive techniques, our results are not as good, as our system struggles with the detection of more complex techniques, especially those less represented in the dataset.

Due to time constraints, we were unable of performing multiple techniques setups. We believe that with further exploration in this direction, better results can be obtained.

For future work, we plan to explore different architectures and transfer learning techniques in order to further improving the performance of our systems. We plan to use the data provided in the competition for the task of media bias detection as we have already proved that the transfer of knowledge from one task to the other is beneficial.

Acknowledgments

This work was supported by the Spanish Ministry of Science and Innovation under the research project FAIRTRANSNLP-DIAGNÓSTICO: Midiendo y cuantificando el sesgo y la justicia en sistemas de PLN (PID2021-124361OB-C32). This work has been also funded by the Ministry of Universities and the European Union through the EuropeaNextGenerationUE funds and the "Plan de Recuperación, Transformación y Resiliencia".

References

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020. We can detect your bias: Predicting the political ideology of news articles. *arXiv preprint arXiv:2010.05338*.

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. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Christoph Hube and Besnik Fetahu. 2018. Detecting biased statements in wikipedia. In *Companion proceedings of the the web conference 2018*, pages 1779–1786.

Fang Jia. 2020. Misinformation literature review: Definitions, taxonomy, and models. *International Journal of Social Science and Education Research*, 3(12):85–90.

Konstantina Lazaridou and Ralf Krestel. 2016. Identifying political bias in news articles. *Bulletin of the IEEE TCDL*, 12.

Xiaoya Li, Xiaofei Sun, Yuxian Meng, Junjun Liang, Fei Wu, and Jiwei Li. 2019. Dice loss for data-imbalanced nlp tasks. *arXiv preprint arXiv:1911.02855*.

G Martino, Alberto Barrón-Cedeno, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020a. Semeval-2020 task 11: Detection of propaganda techniques in news articles. *arXiv preprint arXiv:2009.02696*.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020b. A survey on computational propaganda detection. *arXiv preprint arXiv:2007.08024*.

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 multi-lingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation*, SemEval 2023, Toronto, Canada.

Ruoxi Qin, Kai Qiao, Linyuan Wang, Lei Zeng, Jian Chen, and Bin Yan. 2018. Weighted focal loss: An effective loss function to overcome unbalance problem of chest x-ray14. In *IOP Conference Series: Materials Science and Engineering*, volume 428, page 012022. IOP Publishing.

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.

Francisco-Javier Rodrigo-Ginés, Jorge Carrillo-de Albornoz, and Laura Plaza. under review. A systematic review on media bias detection: what is media bias, how it is expressed, and how to detect it. *Computer Science Review*.

Timo Spinde, Manuel Plank, Jan-David Krieger, Terry Ruas, Bela Gipp, and Akiko Aizawa. 2022. Neural media bias detection using distant supervision with babe–bias annotations by experts. *arXiv preprint arXiv:2209.14557*.

Timo Spinde, Lada Rudnitckaia, Kanishka Sinha, Felix Hamborg, Bela Gipp, and Karsten Donnay. 2021. Mbic–a media bias annotation dataset including annotator characteristics. *arXiv preprint arXiv:2105.11910*.

Taha ValizadehAslani, Yiwen Shi, Jing Wang, Ping Ren, Yi Zhang, Meng Hu, Liang Zhao, and Hualou Liang. 2022. Two-stage fine-tuning: A novel strategy for learning class-imbalanced data. *arXiv preprint arXiv:2207.10858*.

Shehel Yoosuf and Yin Yang. 2019. Fine-grained propaganda detection with fine-tuned bert. In *Proceedings of the second workshop on natural language processing for internet freedom: censorship, disinformation, and propaganda*, pages 87–91.

A Teams results

Subtask 1 - Language: English		
Ranking	**Team**	**Macro F1-score**
1	MELODI	0.78431
2	MLModeler5	0.61632
8	**UnedMediaBiasTeam**	**0.52361**
16	Baseline	0.28802
22	ssnNlp	0.00000

Subtask 3 - Language: English		
Ranking	**Team**	**Micro F1-score**
1	APatt	0.37562
2	vera	0.36802
18	**UnedMediaBiasTeam**	**0.24070**
19	Baseline	0.19517
23	kb	0.06022

Subtask 1 - Language: Italian		
Ranking	**Team**	**Macro F1-score**
1	Hitachi	0.76832
2	QUST	0.76680
5	**UnedMediaBiasTeam**	**0.58408**
12	Baseline	0.38940
17	E8IJS	0.12146

Subtask 3 - Language: Italian		
Ranking	**Team**	**Micro F1-score**
1	KInIT	0.55019
2	NAP	0.53879
16	Baseline	0.39719
17	**UnedMediaBiasTeam**	**0.31717**
20	SinaaAI	0.20284

Subtask 1 - Language: Russian		
Ranking	**Team**	**Macro F1-score**
1	Hitachi	0.75494
2	vera	0.72871
12	Baseline	0.39831
13	**UnedMediaBiasTeam**	**0.36457**
17	E8IJS	0.17460

Subtask 3 - Language: Russian		
Ranking	**Team**	**Micro F1-score**
1	KInIT	0.38682
2	TeamAmpa	0.37781
15	Baseline	0.20722
17	**UnedMediaBiasTeam**	**0.18304**
19	QUST	0.10048

Subtask 1 - Language: French		
Ranking	**Team**	**Macro F1-score**
1	UMUTeam	0.83547
2	QCRITeam	0.76744
10	Baseline	0.56806
11	**UnedMediaBiasTeam**	**0.46536**
17	E8IJS	0.08000

Subtask 3 - Language: French		
Ranking	**Team**	**Micro F1-score**
1	NAP	0.46869
2	TeamAmpa	0.43442
16	Baseline	0.24014
18	**UnedMediaBiasTeam**	**0.23590**
20	SinaaAI	0.19523

Subtask 1 - Language: German		
Ranking	**Team**	**Macro F1-score**
1	UMUTeam	0.81951
1	vera	0.81951
8	Baseline	0.62963
13	**UnedMediaBiasTeam**	**0.36203**
16	MELODI	0.00000

Subtask 3 - Language: German		
Ranking	**Team**	**Micro F1-score**
1	KInIT	0.51304
2	NAP	0.50953
16	**UnedMediaBiasTeam**	**0.31827**
17	Baseline	0.31667
20	SinaaAI	0.04208

Subtask 1 - Language: Polish		
Ranking	**Team**	**Macro F1-score**
1	SharoffAndLepekhin	0.78551
2	Hitachi	0.77922
11	**UnedMediaBiasTeam**	**0.50700**
12	Baseline	0.48962
17	MELODI	0.00000

Subtask 3 - Language: Polish		
Ranking	**Team**	**Micro F1-score**
1	KInIT	0.43037
2	NAP	0.42180
15	**UnedMediaBiasTeam**	**0.23652**
18	Baseline	0.17928
20	SinaaAI	0.06370

Subtask 1 - Language: Spanish		
Ranking	**Team**	**Macro F1-score**
1	DSHacker	0.56349
2	QUST	0.55236
9	**UnedMediaBiasTeam**	**0.33614**
15	MaChAmp	0.21212
16	Baseline	0.15385

Subtask 3 - Language: Spanish		
Ranking	**Team**	**Micro F1-score**
1	TeamAmpa	0.38106
2	KInIT	0.38035
11	Baseline	0.24843
13	**UnedMediaBiasTeam**	**0.22686**
17	QUST	0.12617

Subtask 1 - Language: Greek		
Ranking	**Team**	**Macro F1-score**
1	SinaaAI	0.80588
2	UMUTeam	0.76700
9	**UnedMediaBiasTeam**	**0.52128**
15	Baseline	0.17054
16	E8IJS	0.05670

Subtask 3 - Language: Greek		
Ranking	**Team**	**Micro F1-score**
1	KInIT	0.26733
2	QCRITeam	0.26481
13	**UnedMediaBiasTeam**	**0.10566**
14	Baseline	0.08831
16	SATLab	0.00000

Subtask 1 - Language: Georgian		
Ranking	**Team**	**Macro F1-score**
1	Riga	1.00000
2	vera	0.96268
10	**UnedMediaBiasTeam**	**0.48630**
13	Baseline	0.25641
15	E8IJS	0.00000

Subtask 3 - Language: Georgian		
Ranking	**Team**	**Micro F1-score**
1	KInIT	0.45714
2	QCRITeam	0.41353
11	**UnedMediaBiasTeam**	**0.18012**
14	Baseline	0.13793
16	SATLab	0.07568