Appeal for attention at SemEval-2023 Task 3: Data augmentation and extension strategies for detection of online news persuasion techniques

Amihaesei Sergiu **Cornei Laura** **Stoica George**
Faculty of Computer Science
"Alexandru Ioan Cuza" University of Iasi, Romania
{amihaeseisergiu,cornei.laura10,sgeorge.sstoica99}@gmail.com

Abstract

In this paper, we proposed and explored the impact of four different dataset augmentation and extension strategies that we used for solving the subtask 3 of SemEval-2023 Task 3: multi-label persuasion techniques classification in a multi-lingual context. We consider two types of augmentation methods (one based on a modified version of synonym replacement and one based on translations) and two ways of extending the training dataset (using filtered data generated by GPT-3 and using a dataset from a previous competition). We studied the effects of the aforementioned techniques by using the augmented and/or extended training dataset to fine-tune a pretrained XLM-RoBERTa-Large model. Using the augmentation methods alone, we managed to obtain 3^{rd} place for English, 13^{th} place for Italian and between the 5^{th} to 9^{th} places for the other 7 languages during the competition.

1 Introduction

The subtask 3 of SemEval-2023 Task 3 (Piskorski et al., 2023) focuses on detecting the persuasion techniques used in online news, in a multilingual context. This task is significantly important as it may enable one to identify attempts of manipulation and misinformation, attacks on reputation, persuasive language and logical fallacies in textual data.

The data offered by the organizers of the competition covered nine languages: English, French, Italian, Spanish, German, Georgian, Greek, Polish and Russian. For three of the languages (Spanish, Georgian and Greek) only the test data was available, to promote the development of language-agnostic solutions in a zero-shot learning scenario.

To solve this subtask, we used a pre-trained multi-lingual transformer-based model: XLM-RoBERTa-Large (Conneau et al., 2020). We further fine-tuned this model on an augmented version of the dataset which encompassed the provided training data for the six languages.

We used two augmentation techniques that had the purpose to balance the amount of data associated either with each persuasion technique or with each language. The first augmentation method consisted in creating new articles by translating the data from one language to all of the other languages using a translation API. The second technique involved employing a special procedure to balance the data corresponding to each persuasion technique; in this case, the augmentation consisted in creating similar articles by replacing words from the paragraphs with synonyms from the NLTK WordNet corpus [1].

By using data augmentation, we managed to score 3^{rd} on the English language, and between $5-13^{th}$ places on all the other languages on the official SemEval-2023 test set leaderboard.

Post-competition, we considered extending the training dataset in two ways: by using a dataset from a similar previous competition (Dimitrov et al., 2021) and by using data generated with GPT-3 (Brown et al., 2020). We performed experiments with all four techniques of dataset extension and augmentation and summarized which methods determined our model to obtain the highest F1 score for each language, providing also a comparative analysis with our results during the competition.

While most novel approaches focus on the architecture of the used model(s), for this subtask we tried to emphasize the importance of using augmented and/or additional training data in influencing a model's predictions and performance.

The complete code for our proposed solution, including the data generation and augmentation steps, can be found on Github [2] .

[1] https://www.nltk.org/howto/wordnet.html
[2] https://github.com/Amihaeseisergiu/Appeal-for-Attention-SemEval2023

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 616–623
July 13-14, 2023 ©2023 Association for Computational Linguistics

2 Background

2.1 Related work

To set a starting foundation for our work, we analyzed the state of the art techniques presented at the SemEval's previous proceedings on similar (sub)tasks: the detection of persuasion / propaganda techniques in textual datasets (Da San Martino et al., 2020), (Dimitrov et al., 2021). It is important to mention that the provided data for these previous competitions was exclusively in English.

We observed that the most performant solutions were ensemble based, combining multiple types of pretrained transformers: (Tian et al., 2021), (Abujaber et al., 2021), or combining fine-tuned transformer models with classical Machine Learning models, such as logistic regression and decision trees in the case of (Hossain et al., 2021).

We discovered that many contestants (Abujaber et al., 2021) (Gupta and Sharma, 2021) employed augmentation techniques to balance the given datasets and to improve the performance of their models.

Back-translation was one of the most commonly utilized augmentation methods to enrich the provided monolingual datasets. Authors of (Abujaber et al., 2021) chose as intermediate languages for translation Dutch and Russian, while German and Russian was used in (Ghadery et al., 2021). In (Gupta and Sharma, 2021) it was stated that repeated back translation could have produced many similar data instances, hindering the learning process for the complex proposed models.

EDA (Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks) (Wei and Zou, 2019) was another popular utilized augmentation method, encompassing a series of four operations: synonym replacement (SR), random insertion (RI), random swap (RS) and random deletion (RS). However, the authors introducing EDA mentioned that the techniques might bring negligible improvements when using pre-trained models (including transformer based models such as BERT (Devlin et al., 2019)). In (Gupta and Sharma, 2021), the authors explicitly stated that EDA brought no significant improvements for the task of detecting persuasion techniques in texts.

Starting from these observations, we decided to tackle this year's current task by experimenting with various combinations of data augmentation and data extension techniques that might increase the performance of the transformer-based models. Analyzing the results from the paper (Bayer et al., 2022), we decided to try two augmentation techniques: synonym replacement based on part-of-speech tags and translations from one language to all the other languages. We also considered extending the provided dataset with data generated and annotated by GPT-3 (Ding et al., 2022) and with the dataset from a similar previous competition (Dimitrov et al., 2021).

2.2 Competition dataset

The provided dataset for the subtask 3 was represented by a list of articles in nine languages (English, French, German, Italian, Polish, Russian, Spanish, Georgian and Greek), given as individual files. For the last three languages only the test data was available.

Each article had a unique identifier and was composed of multiple paragraphs. The persuasion techniques (labels) were given in a separate file, the labels being linked to the article id and to the number of the paragraph.

The data was collected from 2020 to mid 2022, and included widely-discussed topics such as COVID-19, climate change, the Russo-Ukrainian war, etc. The labels were not balanced across the languages, with some being entirely absent. Furthermore, the dataset was noisy, for example in some cases the first paragraph of the article was included in the title.

3 System overview

3.1 Data augmentation and extension

To increase our training data and balance the number of paragraphs with an associated language and with associated persuasion techniques, we considered multiple options for modifying the provided competition dataset:

- Extending the competition training dataset with the textual dataset from SemEval 2021 - Task 6 (Dimitrov et al., 2021).
- Extending the competition training dataset with synthetic data generated with GPT-3.
- Augmenting the competition training dataset by adding translations of the articles from each language to all of the other languages.
- Augmenting the competition training dataset using synonyms from WordNet.

We performed multiple experiments in which we applied, in the above mentioned order, a subset of these dataset extension and/or augmentation techniques.

3.1.1 Dataset from SemEval 2021 -Task 6

The SemEval 2021 - Task 6 challenge involved the detection of persuasion techniques in memes and texts of the memes. The task was monolingual, covering only the English language. We selected from the textual dataset provided for this competition only those instances labelled with persuasion techniques that matched the propaganda techniques given in our task.

3.1.2 Dataset generated with GPT-3

We used the GPT-3 Text Davinci model from Open AI [3] to automatically create paragraphs containing a subset of the given propaganda techniques with a maximum length of 500 tokens. We generated 500 paragraphs for each of the 6 languages: English, French, German, Italian, Polish, Russian, by using a customizable template: "Generate a paragraph containing $< persuasion\_techniques >$ as persuasion techniques in $< language >$".

The number of propaganda techniques included for each paragraph was a random number between 1 and 6. The probability of choosing one of the six propaganda techniques with the highest frequencies (Appeal to Fear - Prejudice, Loaded Language, Exaggeration-Minimisation, Name Calling - Labeling, Doubt, Questioning the Reputation) was smaller (0.0166) than the probability of choosing one of the other propaganda techniques (0.05295).

We filtered the results to avoid cases in which GPT-3 would generate definitions of propaganda techniques and not examples of usage. The filtering was done by searching for words with more than 5 characters, that were part of the names of the propaganda techniques. If such words were encountered then the paragraph was not considered in the final dataset.

3.2 Translation based augmentation

To increase the amount of training data and balance the number of paragraphs associated to each language, we employed a translation based augmentation technique. We used the DeepTranslator API [4] to translate each article from the dataset from the original language to all of the other 8 languages.

[3]https://platform.openai.com/docs/models/gpt-3

[4]https://pypi.org/project/deep-translator/

3.3 Synonym based augmentation

In order to further extend our dataset and to balance the number of paragraphs corresponding to each persuasion technique, we used a synonym-based augmentation technique. More exactly, we considered the set of paragraphs that had associated at least one of the less represented labels (with a frequency smaller than 500) and performed an oversampling over it (with a sampling factor of 1.5). For each paragraph in the set we created a new one by replacing every word in it, if possible, with a WordNet synonym. Each synonym was chosen such that it had the same part of speech as the original word and a Levenshtein distance [5] between it and the original word of at least 3. We used the Spacy library [6] to retrieve the parts of speech, and the NLTK WordNet corpus to extract matching synonyms. We finally extended our dataset with the newly created paragraphs.

3.4 Data preprocessing and used model

To solve subtask 3, we utilized the XLM-RoBERTa-Large pre-trained multi-lingual model (Conneau et al., 2020), fine-tuning it on the augmented and/or extended training dataset.

In the tests performed during the competition, we utilized the training data augmented using the two previously discussed techniques (synonym based augmentation and translation based augmentation). In the experiments realized post-competition, we considered multiple ways of combining the four augmentation and extension methods mentioned in section 3.1.

Before feeding the data to the model, we preprocessed each paragraph using the common configuration for RoBERTa (Zhuang et al., 2021), which allows a maximum sequence length of 512.

4 Experimental setup

In our experiments, we utilized the Google Colab[7] and Kaggle[8] platforms to fine-tune our XLM-RoBERTa-Large model. We cumulated the given train and dev splits, using 80% of it for training and 20% for validation.

Due to the dimensions of our model and the limited memory in our used environments, we performed a manual hyper-parameter tuning. We se-

[5]https://pypi.org/project/python-Levenshtein/

[6]https://spacy.io/

[7]https://colab.research.google.com/

[8]https://www.kaggle.com/

lected as final hyperparameters the ones that enabled our model to obtain the highest validation F1 Micro score.

To fine-tune the model on this multi-label classification task we used the $MultiLabelClassificationModel$ class provided by the Simple Transformers library [9]. The official evaluation metric used for this subtask was the Micro F1 score; however, the Macro F1 score was also computed.

5 Results

5.1 Competition results

As previously mentioned in section 3.4, the final used dataset during the competition was an augmented version of the original dataset. Table 1 contains the hyperparameters used to fine-tune the XLM-RoBERTa-Large model: a learning rate of $1e^{-5}$, a weight decay of $5e^{-10}$ and a number of training epochs equal to 20; we also used a batch size of 16 with 16 gradient accumulation steps

Training parameter	Value
learning_rate	1e-5
num_train_epochs	20
weight_decay	5e-10
train_batch_size	16
gradient_accumulation_steps	16
eval_batch_size	16

Table 1: Hyperparameters used during competition

Table 2 highlights the results obtained by our team (ACS), in comparison to the scores attained by the team with rank 1 and by the baseline solution.

We obtained 3^{rd} place on the English language, with a Micro F1 score of 0.36299, very close to the score corresponding to the 1^{st} place: 0.37562. We also obtained 13^{th} place for Italian and ranks between 5 and 9 for the other seven languages.

We believe that some reasons justifying the good position on the English language might include:

- There are a lot more available corpora in English than in most other languages. Authors of the XLM-RoBERTa-Large (Conneau et al., 2020) stated that the used amount of data for training the model was considerably higher for the English language than for the other languages.
- Regarding the augmentation step, the WordNet version for English is more developed in comparison to the versions for the other languages. The large amount of English resources might also contribute to a higher quality of services such as part-of-speech tagging using Spacy or translations using DeepTranslator.

Lg.	Rank	Team	F1 Micro	F1 Macro
English	1	APatt	0.37562	0.12919
	...	...	...	...
	3	**ACS**	**0.36299**	**0.16621**
	...	...	...	...
	19	Baseline	0.19517	0.06925
Italian	1	KInIT	0.55019	0.21436
	...	...	...	...
	13	**ACS**	**0.43097**	**0.21096**
	...	...	...	...
	16	Baseline	0.39719	0.12152
Russian	1	KInIT	0.38682	0.18879
	...	...	...	...
	6	**ACS**	**0.31152**	**0.17325**
	...	...	...	...
	15	Baseline	0.20722	0.08598
French	1	NAP	0.46869	0.32171
	...	...	...	...
	9	**ACS**	**0.37417**	**0.20318**
	...	...	...	...
	16	Baseline	0.24014	0.09867
German	1	KInIT	0.51304	0.23313
	...	...	...	...
	8	**ACS**	**0.41783**	**0.21802**
	...	...	...	...
	17	Baseline	0.31667	0.08345
Polish	1	KInIT	0.43037	0.17929
	...	...	...	...
	8	**ACS**	**0.34368**	**0.20072**
	...	...	...	...
	18	Baseline	0.17928	0.05932
Spanish	1	TeamAmpa	0.38106	0.24366
	...	...	...	...
	5	**ACS**	**0.31730**	**0.13934**
	...	...	...	...
	11	Baseline	0.24843	0.02007
Greek	1	KInIT	0.26733	0.12559
	...	...	...	...
	6	**ACS**	**0.20583**	**0.11851**
	...	...	...	...
	14	Baseline	0.08831	0.00606
Georgian	1	KInIT	0.45714	0.32758
	...	...	...	...
	7	**ACS**	**0.27986**	**0.26142**
	...	...	...	...
	14	Baseline	0.13793	0.14083

Table 2: Competition results of our team (ACS)

5.2 Post competition results

Post-competition, we performed multiple experiments to evaluate the impact of the four augmentation and extension techniques proposed in section 3.1. We were interested to see which combinations

[9]https://simpletransformers.ai/

of these techniques produced the highest increase in the Micro F1 score for each of the 9 languages. We continued to utilize the same model: XLM-RoBERTa-Large, as during the competition.

For easier reference, we abbreviated:

- the synonym based augmentation to s
- the translation based augmentation to t
- the extension made using data generated with GPT-3 to g
- the extension realized using the dataset from task 6 of the previous year's competition to o.

To limit the number of possibilities, we considered a fixed order of application of the methods: o, g, t, s. Next, we created all of the 16 datasets by applying none, one, or multiple of the four techniques over the original training data.

We fine-tuned the model on each generated training dataset and registered the Micro F1 scores obtained on the test dataset for every language.

Regarding hyperparameter tuning, we observed that the only parameter generating a substantial modification of the Micro F1 score was the number of epochs (denoted by E). Therefore, we decided to work with the same values for the hyperparameters as the ones used during the competition (Table 1) and varied only the number of epochs.

Table 3 summarizes for each language the used number of epochs (E) and the subset of utilized augmentations and/or extension methods (M) that generated the maximum Micro F1 score on the test dataset. The last column of the table highlights the increase in score in comparison to the results obtained during the competition.

Lang.	E	F1 Micro	M	Increase
English	20	0.38628	t	0.02329
Italian	25	0.48011	o, g, s	0.04914
Russian	25	0.33136	o, g, s	0.01984
French	25	0.39364	o, g, s	0.01947
German	20	0.46133	s	0.0435
Polish	20	0.37195	g, t, s	0.02827
Spanish	20	0.31730	t, s	0
Greek	20	0.23529	o	0.02946
Georgian	20	0.33544	g, t	0.05558

Table 3: Post-competition results emphasizing the number of epochs and the techniques used to obtain the highest F1 score for each language

Analyzing table 3, it can be observed that the highest increase in the F1 score is obtained for the Georgian, Italian and German languages, followed by Polish and Greek. The subset of augmentation and/or extension techniques that generated the best results varied greatly from language to language.

Appendix A contains the number of articles labelled with each persuasion technique, resulted after applying different combinations of the used data augmentation and/or extension techniques. The results are cumulated for all languages.

Appendix B indicates the detailed results obtained post competition for 10 subsets of augmentation and/or extension methods, for which we obtained the highest mean over the Micro F1 scores obtained on the validation dataset for all languages. The third column in the associated table indicates the Micro F1 scores attained on the test dataset.

6 Conclusion

In this paper, we propose and study the effect of four data augmentation and extension techniques to solve the problem of identifying persuasion techniques used in online news in a multi-lingual setup (subtask 3 of Task 3 from the SemEval 2023 competition).

In case of augmentation, we utilize two methods: a variation of the classic synonym replacement operation that takes into account the words' part of speech and Levenshtein distances and a translation based techniques that creates new articles by translating the data from one language to all of the other languages. We also consider extending the training dataset with data from a similar previous SemEval task and with synthetic data generated with GPT-3.

We explore the impact of these techniques by using the augmented and/or extended training dataset to fine-tune a pretrained XLM-RoBERTa-Large model.

We manage to achieve good results on all 9 languages proposed in the competition. Our results indicate that the quality and quantity of the training data have a major impact on a model's performance and that we can obtain major improvements by using suitable augmentation and extension techniques.

Acknowledgements

We would like to thank our faculty teachers: Prof. Madalina Raschip, Prof. Mihaela Breaban and Prof. Diana Trandabat for providing valuable knowledge and insights during the Deep Learning with applications in NLP and the Machine Translation classes.

References

Dia Abujaber, Ahmed Qarqaz, and Malak A Abdullah. 2021. LeCun at SemEval-2021 Task 6: Detecting persuasion techniques in text using ensembled pretrained transformers and data augmentation. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 1068–1074.

Markus Bayer, Marc-André Kaufhold, and Christian Reuter. 2022. A survey on data augmentation for text classification. *ACM Computing Surveys*, 55(7):1–39.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

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.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021. SemEval-2021 task 6: Detection of persuasion techniques in texts and images. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 70–98, Online. Association for Computational Linguistics.

Bosheng Ding, Chengwei Qin, Linlin Liu, Lidong Bing, Shafiq Joty, and Boyang Li. 2022. Is gpt-3 a good data annotator? *arXiv preprint arXiv:2212.10450*.

Erfan Ghadery, Damien Sileo, and Marie-Francine Moens. 2021. LIIR at SemEval-2021 task 6: Detection of persuasion techniques in texts and images using CLIP features. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 1015–1019, Online. Association for Computational Linguistics.

Vansh Gupta and Raksha Sharma. 2021. Nlpiitr at semeval-2021 task 6: Roberta model with data augmentation for persuasion techniques detection. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 1061–1067.

Tashin Hossain, Jannatun Naim, Fareen Tasneem, Radiathun Tasnia, and Abu Nowshed Chy. 2021. CSECU-DSG 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, Online. Association for Computational Linguistics.

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.

Junfeng Tian, Min Gui, Chenliang Li, Ming Yan, and Wenming Xiao. 2021. Mind at semeval-2021 task 6: Propaganda detection using transfer learning and multimodal fusion. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 1082–1087.

Jason Wei and Kai Zou. 2019. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6382–6388, Hong Kong, China. Association for Computational Linguistics.

Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. 2021. A robustly optimized bert pre-training approach with post-training. In *Proceedings of the 20th chinese national conference on computational linguistics*, pages 1218–1227.

A Number of articles per each persuasion technique after applying the data augmentation and/or extension techniques

Persuasion technique	none	g	o	t	s	g,o	g,t	g,s	o,t	o,s	t,s	g,o,t	g,o,s	g,t,s	o,t,s	g,o,t,s
Appeal to Authority	746	1285	781	1221	1169	1320	1760	1285	1256	1189	1395	1795	1320	1760	1438	1795
Appeal to Popularity	345	922	345	500	1376	922	1077	922	500	1375	597	1077	922	1077	602	1077
Appeal to Values	646	1137	646	941	1118	1137	1432	1137	941	1086	1198	1432	1137	1432	1196	1432
Appeal to Fear-Prejudice	1528	1692	1619	2388	2431	1783	2552	1692	2479	2578	2747	2643	1783	2552	2857	2643
Flag Waving	708	1164	763	1063	1019	1219	1519	1164	1118	1081	1226	1574	1219	1519	1295	1574
Causal Oversimplification	583	1136	619	928	1002	1172	1481	1136	964	1021	1125	1517	1172	1481	1177	1517
False Dilemma-No Choice	469	964	469	729	1906	964	1224	964	729	1886	853	1224	964	1224	854	1224
Consequential Oversimplification	365	897	365	525	1536	897	1057	897	525	1534	709	1057	897	1057	701	1057
Straw Man	313	771	353	458	1326	811	916	771	498	1478	1686	956	811	916	1826	956
Red Herring	210	778	217	300	872	785	868	778	307	891	1126	875	785	868	1139	875
Whataboutism	149	643	216	249	625	710	743	643	316	878	946	810	710	743	1197	810
Slogans	683	1197	753	1083	993	1267	1597	1197	1153	1055	1229	1667	1267	1597	1313	1667
Appeal to Time	163	783	163	248	698	783	868	783	248	677	890	868	783	868	890	868
Conversation Killer	939	1459	939	1474	1737	1459	1994	1459	1474	1649	1840	1994	1459	1994	1858	1994
Loaded Language	6198	6360	6690	9758	8832	6852	9920	6360	10250	9530	11324	10412	6852	9920	12015	10412
Repetition	959	1534	973	1404	1268	1548	1979	1534	1418	1291	1512	1993	1548	1979	1548	1993
Exaggeration-Minimisation	1613	1806	1712	2623	2516	1905	2816	1806	2722	2617	3062	2915	1905	2816	3177	2915
Obfuscation-Vagueness-Confusion	365	909	372	555	1499	916	1099	909	562	1521	726	1106	916	1099	731	1106
Name Calling-Labeling	4376	4570	4723	6961	6265	4917	7155	4570	7308	6687	8152	7502	4917	7155	8569	7502
Doubt	3684	3864	3795	5789	5085	3975	5969	3864	5900	5249	6521	6080	3975	5969	6683	6080
Guilt by Association	597	1057	597	932	974	1057	1392	1057	932	981	1143	1392	1057	1392	1144	1392
Appeal to Hypocrisy	878	1417	878	1318	1258	1417	1857	1417	1318	1265	1541	1857	1417	1857	1541	1857
Questioning the Reputation	1948	2077	1971	3043	3148	2100	3172	2077	3066	3132	3802	3195	2100	3172	3772	3195

B Post-competition detailed results

Aug.	Mean F1 Micro	Final F1 Micro
g	0.53527	En: 0.33959 It: 0.47421 Ru: 0.31037 Po: 0.35132 Fr: 0.38501 Ge: 0.45958 Es: 0.30643 Gr: 0.22472 Ka: 0.31293
o	0.4756	En: 0.33635 It: 0.46291 Ru: 0.31143 Po: 0.33494 Fr: 0.37467 Ge: 0.42945 Es: 0.31520 Gr: 0.23529 Ka: 0.33457
t	0.638	En: 0.38628 It: 0.46133 Ru: 0.31281 Po: 0.33270 Fr: 0.37564 Ge: 0.43920 Es: 0.31162 Gr: 0.20985 Ka: 0.30389
s	0.677	En: 0.34596 It: 0.47407 Ru: 0.28699 Po: 0.30968 Fr: 0.37520 Ge: 0.46133 Es: 0.30575 Gr: 0.19625 Ka: 0.31452
g, o, t, s	0.658	En: 0.37201 It: 0.46497 Ru: 0.31157 Po: 0.33944 Fr: 0.37389 Ge: 0.42269 Es: 0.31604 Gr: 0.21059 Ka: 0.30619

Aug.	Mean F1 Micro	Final F1 Micro
t, s	0.7266	En: 0.36299 It: 0.43097 Ru: 0.31152 Po: 0.34368 Fr: 0.37417 Ge: 0.41783 Es: 0.31730 Gr: 0.20583 Ka: 0.27986
g, t	0.66185	En: 0.37160 It: 0.44695 Ru: 0.29921 Po: 0.35172 Fr: 0.37674 Ge: 0.43039 Es: 0.30864 Gr: 0.21967 Ka: 0.33544
g, o, s	0.531	En: 0.34934 It: 0.48011 Ru: 0.33136 Fr: 0.39364 Ge: 0.45054 Po: 0.34555 Es: 0.29499 Gr: 0.22452 Ka: 0.31915
g, o, t	0.666	En: 0.36849 It: 0.45310 Ru: 0.29810 Fr: 0.38189 Ge: 0.44043 Po: 0.34976 Es: 0.30559 Gr: 0.21540 Ka: 0.29814
o, t, s	0.72397	En: 0.33959 It: 0.47421 Ru: 0.31037 Fr: 0.38501 Ge: 0.45958 Po: 0.35132 Es: 0.30643 Gr: 0.22472 Ka: 0.31293