SlavicNLP 2025 Shared Task: Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media

Jakub Piskorski,[1] **Dimitar Dimitrov,**[2] **Filip Dobranić,**[3] **Marina Ernst,**[4]
Jacek Haneczok,[5] **Ivan Koychev,**[2] **Nikola Ljubešić,**[6,3] **Michał Marcińczuk,**[7]
Arkadiusz Modzelewski,[8,9] **Ivo Moravski,**[3] **Roman Yangarber**[10]
[1]Institute of Computer Science, Polish Academy of Science, Poland jpiskorski@gmail.com
[2] Sofia University "St. Kliment Ohridski", Bulgaria {ilijanovd, koychev, moravski}@fmi.uni-sofia.bg
[3] Institute for Contemporary History, Ljubljana, Slovenia filip.dobranic@inz.si
[4] University of Koblenz, Germany marinaernst@uni-koblenz.de
[5] Visa Technology Europe jacek.haneczok@gmail.com
[6] Jožef Stefan Institute, Ljubljana, Slovenia nikola.ljubesic@ijs.si
[7] CodeNLP, Poland marcinczuk@gmail.com
[8] University of Padua, Italy arkadiusz.modzelewski@unipd.it
[9] Polish-Japanese Academy of Information Technology, Poland arkadiusz.modzelewski@pja.edu.pl
[10] University of Helsinki first.last@helsinki.fi

Abstract

We present SlavicNLP 2025 Shared Task on *Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media*. The task is structured into two subtasks: (1) *Detection*, to determine whether a given text fragment contains persuasion techniques, and (2) *Classification*, to determine for a given text fragment which persuasion techniques are present therein using a taxonomy of 25 persuasion technique taxonomy. The task focuses on two text genres, namely, parliamentary debates revolving around widely discussed topics, and social media, in five languages: Bulgarian, Croatian, Polish, Russian and Slovene. This task contributes to the broader effort of detecting and understanding manipulative attempts in various contexts. There were 15 teams that registered to participate in the task, of which 9 teams submitted a total of circa 220 system responses and described their approaches in 9 system description papers.

1 Introduction

Persuasion techniques are psychological instruments that people use to influence others' opinions and actions. Some of such techniques use invalid or otherwise faulty reasoning in the construction of an argument, while others intentionally appeal to emotions to cause the recipient of the information to experience certain feelings, e.g. fear, in order to win an argument, especially in the absence of factual evidence.

Persuasion constitutes an essential part of political debates and impacts the outcome of policy-related decisions. Persuasion is also a weapon used by social media influencers to manipulate public opinion. Several shared tasks have been held over the years to study the detection and categorization of persuasive techniques in different text genre and discourse. In this paper, we present SlavicNLP 2025 Shared Task on *Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media*, which uses a taxonomy of 25 fine-grained persuasion techniques and covers 5 Slavic languages, namely, Bulgarian, Croatian, Polish, Russian and Slovene. This task contributes to the broader effort of detecting and understanding influencing and manipulative attempts in parliamentary and social media contexts. 15 teams registered to participate in the task, of which 9 teams submitted a total of circa 220 system responses, and described their approaches in 9 system description papers.

The paper is organized as follows. Section 2 introduces the two subtasks. Section 3 surveys related work. Section 4 describes the training and test datasets created for the task. Section 5 gives an overview of the evaluation framework. Section 6 presents the results of the competition and comparison of the participant systems. Section 7 concludes with a summary of the task.

The views and opinions expressed in this article are solely those of the authors and do not necessarily reflect the official policy or position of Visa. Any statements, insights, or conclusions presented are made in a personal capacity and should not be attributed to Visa or its affiliates.

Proceedings of the 10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025), pages 254–275

2 The Tasks

The task focuses on the detection and classification of Persuasion Techniques in 5 Slavic languages: Bulgarian (BG) , Croatian (HR), Polish (PL), Slovene (SI) and Russian (RU) in two types of texts: (a) parliamentary debates on highly-debated topics (BG, HR, PL, SI), and (b) social media posts related to the spread of disinformation (RU).

The task consists of two subtasks:

1. **Subtask 1: (Detection)** Given a text and a list of text fragment offsets, determine for each corresponding text fragment whether it contains one or more persuasion techniques from a given taxonomy of persuasion techniques,

2. **Subtask 2: (Classification)** Given a text and a list of text fragment offsets, determine for each span which persuasion techniques are present in it (the set could be empty). The text fragments correspond to paragraphs.

Subtask 1 is a binary classification task, whereas Subtask 2 is a multi-class multi-label classification task.

2.1 Taxonomy

In this task we exploit the taxonomy from SemEval 2023 Task 3 (Piskorski et al., 2023c), which is extended by two new persuasion techniques, namely: *false equivalence*,[1] and *appeal to pity*.[2] The extended taxonomy is shown in Figure 1. Definitions and examples are provided in Appendix A

```
ATTACK ON REPUTATION
- Name Calling or Labelling
- Guilt by Association
- Casting Doubt
- Appeal to Hypocrisy
- Questioning the Reputation

JUSTIFICATION
- Flag Waiving
- Appeal to Authority
- Appeal to Popularity
- Appeal to Fear, Prejudice
- Appeal to Values

DISTRACTION
- Strawman
- Whataboutism
- Red Herring
- Appeal to Pity

SIMPLIFICATION
- Causal Oversimplification
- False Dilemma or No Choice
- Consequential Oversimplification
- False Equivalence

CALL
- Slogans
- Conversation Killer
- Appeal to Time

MANIPULATIVE WORDING
- Loaded Language
- Obfuscation, Intentional Vagueness, Confusion
- Exaggeration or Minimisation
- Repetition
```

Figure 1: Two-tier Persuasion Technique taxonomy.

3 Related Work

Parliamentary debates have been receiving considerable attention in the natural language processing community for several reasons. One is the availability of a significant amount of textual data (Erjavec et al., 2024), sometimes with translations into multiple languages (Koehn, 2005). In addition to textual data, recordings are often available, presenting a great opportunity for building speech and text datasets (Ljubešić et al., 2024) not riddled with privacy or copyright concerns. The metadata on speakers allows for various downstream research directions, such as speaker profiling (Ljubešić and Rupnik, 2002) or political leaning analysis (Evkoski and Pollak, 2023). However, the data also allow for further enrichment, such as political agenda (Sebők et al., 2024), or sentiment (Mochtak et al., 2024) , opening up additional research directions (Abercrombie and Batista-Navarro, 2020).

3.1 Related shared tasks

Several shared tasks have been held over the years to study the detection and categorization of persuasive techniques. The first tasks NLP4IF-2019 Shared Task on Fine-Grained Propaganda Detection (Da San Martino et al., 2019); SemEval-2020 Task 11 on Detection of Persuasion Techniques in News Articles (Da San Martino et al., 2020) focused on the detection of persuasion techniques in text fragments and document-level classification with an initial taxonomy of 18 techniques in English news articles. Later on, SemEval-2023 Task 3 on Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup (Piskorski et al., 2023b) refined and extended the taxonomy to a total of 23 persuasion techniques, grouped in 6 different categories. Moreover, the task introduced texts in nine languages, including Slavic languages like Polish and Russian, enabling multilingual research. Furthermore, CLEF 2024 Task 3 on Persuasion

[1] https://www.logicallyfallacious.com/logicalfallacies/False-Equivalence

[2] https://www.logicallyfallacious.com/logicalfallacies/Appeal-to-Pity

Techniques (Piskorski et al., 2024) built upon the Semeval-2023 Task 3 by including new articles in five languages, two of which from the Slavic family—Arabic, *Bulgarian*, English, Portuguese, and *Slovene*. Unlike the tasks mentioned so far, which focus on news articles, DIPROMATS 2023: automatic detection and characterization of propaganda techniques in messages from diplomats and authorities of world powers (Moral et al., 2023) and DIPROMATS 2024: Detection, characterization and tracking of propaganda in messages from diplomats and authorities of world powers (Moral et al., 2024) turn their attention to diplomatic tweets, releasing a dataset of more than 21,000 tweets in English and Spanish, posted by authorities of China, Russia, United States and the European Union, a novel angle that explores governmental propaganda directly at its source. Additionally, the studies adapt the original (Da San Martino et al., 2019) taxonomy to a new version of 15 persuasion techniques. Other shared tasks on persuasion that build upon the already mentioned taxonomies but concerning content in Arabic include: (Alam et al., 2022; Hasanain et al., 2023)

In parallel to language analysis, several tasks were organized on the detection of persuasion techniques in multimodal content (vision language), particularly in memes SemESemEval-2021 6 on Detection of Persuasion Techniques in Texts and Images (Dimitrov et al., 2021). This shared task featured an extension of the Semeval-2020 taxonomy, incorporating persuasion only found in the visual content, totaling 22 techniques—20 multimodal and 2 vision-only. The presented dataset, collected from Facebook public groups, consisted of 950 English memes. Subsequently, SemEval-2024 Task 4 on Multilingual Detection of Persuasion Techniques in Memes (Dimitrov et al., 2024) significantly increased the data with more than ten thousand memes, with the addition of memes in two Slavic languages—Bulgarian and North Macedonian. Hasanain et al. (2024) also conducted a task on memes but applied it to Arabic multimodal content.

In contrast to existing shared tasks, this edition of SlavicNLP focuses on detecting and classifying persuasion techniques in parliamentary debates from various Slavic-speaking countries, with the aim of improving the understanding of how political leaders influence public opinion, guide policy decisions, and frame key issues. To the best of our knowledge, this is the first shared task that focuses on the domain of parliamentary debates.

	BG	PL	RU	SI
Documents	20	15	27	15
Paragraphs	363	289	239	108
Paragraphs with PTs	168	194	166	58
Text spans annotated	756	886	256	632
PTs covered	25	25	24	23
AVG words/document	1126	1280	327	1164

Table 1: Training data statistics across languages: PT—Persuasion Techniques.

4 Datasets

4.1 Training data

As training data, the task exploits a number of pre-existing datasets with text span-level and paragraph-level annotated persuasion techniques, created for prior SemEval tasks (Da San Martino et al., 2020; Dimitrov et al., 2021; Piskorski et al., 2023b) and CLEF 2024 (Piskorski et al., 2024). For three Slavic languages—Bulgarian, Polish, and Russian—annotated data were available from these resources. Participants were provided with a small additional domain-tailored (i.e., parliamentary debates, social media) training dataset, whose statistics are in Table 1. Note that Croatian training data was not available.

4.2 Test data

The test data set comprises 206 documents in five languages. Most documents are excerpts from parliamentary session transcriptions covering a variety of topics, except for the Russian subset, which consists of news articles. The dataset covers 25 persuasion techniques, with the number of annotations per technique ranging from 59 for *False Equivalence* to 906 for *Loaded Language*. Table 2 provides detailed test data statistics for each language, and Figure 2 provides the distribution of the persuasion techniques and comparison across the languages.

4.2.1 Bulgarian

For Bulgarian, the document captures a deeply polarized debate in the Bulgarian parliament, centered on Bulgaria's foreign policy (especially regarding military aid to Ukraine) against a backdrop of domestic discontent, concerns about national sovereignty and broader questions about Bulgaria's place in international order. The discussions intertwine international issues (Gaza, Western Sahara,

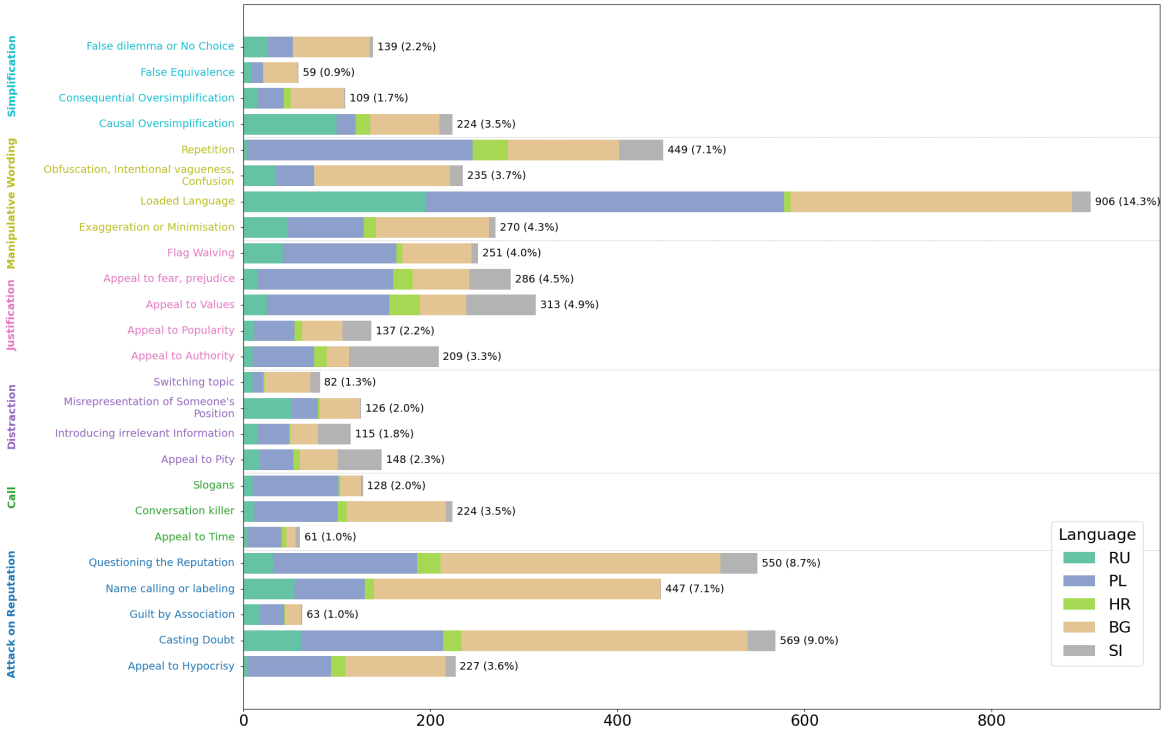

Figure 2: Distribution of persuasion technique annotations (per language)

Ukraine), domestic priorities, procedural integrity, and national identity, reflecting the complexity and intensity of current Bulgarian political discourse.

4.2.2 Polish

For Polish, the documents contain recent debates in the Polish parliament on several major topics: the highly polarized dispute over abortion laws, national security and defense policy, Poland's role and challenges within the European Union, proposed amendments to strengthen hate crime and anti-discrimination laws; and a range of social and economic issues, including vaccination policy, forest management, mass layoffs, mental health awareness, and calls for better wages for school support staff.

4.2.3 Croatian

For Croatian, the documents contain extensive debates from the Croatian Parliament in 2022, focusing on legislative challenges, economic pressures, and social issues. Central topics include the controversial Law on Land Consolidation and its implementation barriers, disputes over constitutional reforms and the role of the Constitutional Court, and the handling of referendums. Economic discussions address energy strategy amid the Ukraine war, inflation, the adoption of the euro and government interventions in energy pricing. The debates also highlight EU-funded regional development, mental health, labor law, food security, the introduction of ecocide as a crime, and persistent concerns about corruption, media independence, and the legacy of historical events and minority rights.

4.2.4 Russian

For Russian, the documents cover a range of current social and political issues. They discuss the dynamics of the Ukraine-Russia war, including perceptions of negotiations between Putin and Trump, and Russia's broader geopolitical struggle with the West. Topics also include the spread of disinformation, particularly how Ukrainian actors exploit Russian officials' weaknesses to undermine trust in government and sow confusion among the public. Significant attention is paid to demographic challenges; migration and integration problems related to tensions between native Russians and migrants from Central Asia, as well as the perceived failure of state policies to foster assimilation or protect Russian interests. The resilience of civilians in conflict zones and the importance of national unity and faith are emphasized as key themes for Russia's future.

4.2.5 Slovene

For Slovene, the documents cover major topics from Slovenian parliamentary sessions, including

	BG	HR	PL	RU	SI
Documents	59	10	38	63	36
Paragraphs	1,361	74	729	590	487
Paragraphs with PTs	735	34	462	444	157
Text spans annotated	2,520	263	2,160	833	551
PTs covered	25	23	25	25	25
AVG words/document	1,153	1,192	1,189	373	1,059

Table 2: Test data statistics across languages

government oversight of police operations and mechanisms for ministerial supervision. They address economic responses to the COVID-19 crisis, notably Slovenia's participation in the Pan-European Guarantee Fund to support businesses. Amendments to the State Administration Act are discussed, focusing on reorganizing ministerial responsibilities, such as forestry and military heritage management. The documents contain debates on social issues like rising poverty and energy prices. The annual report of the Human Rights Ombudsman is reviewed, emphasizing the impact of pandemic measures on human rights and state accountability. Infrastructure and energy policy, including the operation of Maribor Airport and strategies for energy independence, are also covered.

4.3 Annotation process

For the annotation of the documents with persuasion techniques, a dedicated team was set up for each of the five languages of the task, supervised by a designated language coordinator, and consisted of two to four annotators and one curator. Most of the annotators were native speakers and had prior experience in linguistic annotations, in particular, in the area of propaganda and manipulative narratives annotations. The background of annotators covered various disciplines, including, i.a., computational linguistics and humanities and social sciences, some of which were students. The annotators underwent comprehensive training, which involved studying the detailed annotation guidelines (Piskorski et al., 2023a).

Each document was annotated by two annotators. Given the complexity of annotating persuasion techniques in texts (Stefanovitch and Piskorski, 2023) a curator was assigned to each language to verify adherence to predefined guidelines and to systematically review the annotations, assess their accuracy and quality, and merge and select the most appropriate ones. Regular meetings were conducted in each language team, and across languages—to resolve disagreements and maintain consistency in annotations.

For the task of annotating documents with persuasion techniques we adapted INCEpTION (Klie et al., 2018), a web-based collaborative annotation framework.

5 Evaluation Framework

5.1 Evaluation Measures

The following are used as official metrics for ranking the participant systems on the two **subtasks**:

1. **Detection**: F_1
2. **Classification**: macro and micro F_1

For Subtask 2, we also computed F_1 scores for the classification of each type of persuasion technique to compare the results of the shared task participants with the results obtained on similar tasks organized in some recent competitions at SemEval and CLEF.

5.2 Official formats and naming conventions

5.2.1 Source Documents

The files containing the source documents use `UTF-8` encoding and have a name starting with 2-letter encoding of the language (capitalized) and followed by an underscore and a unique identifier, e.g., `PL_article123`.

5.2.2 Gold-label file(s)

For Subtask 1 the gold-label file consist of lines, where each line consists of four tab-delimited elements:

`articleID start end persuasion_flag`

where `persuasion_flag` indicates whether the text fragment starting at `start` and ending at `end` character position in the document `articleID` contains at least one persuasion technique.

For Subtask 2 the gold-label file consists of lines, where each line consists of three or more tab-delimited elements in the following format:

`articleID start end pt1 ... ptN`

where `pt1, ... ptN` is a list of `N` labels (might be empty) corresponding to persuasion techniques present in the text fragment starting at `start` and ending at `end` character position in the document `articleID`.

5.2.3 Submission file(s)

The submission files have a format identical to that of the gold-standard label files described above.

BG		HR		PL		RU		SI	
Team	F_1	Team	F_1	Team	F_1	Team	F_1	Team	F_1
FactUE	0.88	FactUE	0.96	oplot	0.90	INSAntive	0.87	UFAL4DEM	0.86
baseline	0.88	baseline	0.94	syntax_squad	0.90	Gradient-Flush	0.86	FactUE	0.85
oplot	0.87	UFAL4DEM	0.94	FactUE	0.90	UFAL4DEM	0.86	baseline	0.85
syntax_squad	0.87	oplot	0.92	baseline	0.90	FactUE	0.84	oplot	0.85
UFAL4DEM	0.86	INSAntive	0.89	UFAL4DEM	0.89	baseline	0.83	syntax_squad	0.82
Gradient-Flush	0.84	Gradient-Flush	0.85	Gradient-Flush	0.88	oplot	0.83	Gradient-Flush	0.81
PSAL_NLP	0.82	PSAL_NLP	0.83	INSAntive	0.88	syntax_squad	0.80	INSAntive	0.65
INSAntive	0.81			PSAL_NLP	0.83	PSAL_NLP	0.73	PSAL_NLP	0.62

Table 3: Subtask 1: F_1 scores

BG		HR		PL		RU		SI	
Team	F_1^M	Team	F_1^M	Team	F_1^M	Team	F_1^M	Team	F_1^M
PSAL_NLP	0.32	Gradient-Flush	0.36	PSAL_NLP	0.32	PSAL_NLP	0.21	PSAL_NLP	0.26
INSAntive	0.21	UFAL4DEM	0.33	FactUE	0.29	INSAntive	0.18	Gradient-Flush	0.19
UFAL4DEM	0.19	PSAL_NLP	0.32	Gradient-Flush	0.28	Gradient-Flush	0.13	UFAL4DEM	0.15
oplot	0.19	oplot	0.28	INSAntive	0.26	oplot	0.13	baseline	0.14
Gradient-Flush	0.17	baseline	0.21	UFAL4DEM	0.23	UFAL4DEM	0.11	INSAntive	0.14
dutir	0.15	dutir	0.18	oplot	0.21	FactUE	0.02	oplot	0.11
baseline	0.07	INSAntive	0.18	dutir	0.21	baseline	0.01	dutir	0.11
FactUE	0.04	FactUE	0.05	baseline	0.10			FactUE	0.02

Table 4: Subtask 2: Macro-averaged F_1 scores

5.3 Task Organization

The shared task was conducted in two phases:

Development Phase: initially, the participants were provided only with references to existing datasets annotated with persuasion techniques, which covered many languages and text genre. Three of the languages—Bulgarian, Polish and Russian—were covered by these datasets. At a later stage, an additional small *training* dataset covering the domain of parliamentary debates and social media was released to the participants in order to better tailor their solutions for the tasks.

Test Phase: in the second phase, the raw documents of the *test* set (without the gold-standard answers) were released. The participants were given approximately 7 days to submit their final predictions on the *test* set for both subtasks. Participants were allowed to submit up to a maximum of 5 responses per language; the response with the best scores was considered for the official rankings of the team.

A total of 15 teams registered to participate in the task. Nine teams submitted system responses, of which 8 submitted valid responses. Seven teams participated in both tasks, while 1 team participated only in subtask 1, and 1 team participated only in subtask 2. In total, 220 valid system responses were submitted and compared.

Official results for the test phases are available on the web site of the shared task.[3] The repository with the evaluation and conformity scripts is at github.com/jacxhanx/PersuasionNLPTools

6 Participants and Results

This section provides the official results on the two subtasks and a comparison of the approaches used by the participants in terms of models and resources exploited. We also compare the participant systems against a transformer-based baseline system.

6.1 Baselines

The main principle behind the development of our baseline systems is fine-tuning a pretrained multilingual language model using the provided task-specific data. We use the *XLM-RoBERTa-base* (Conneau et al., 2019) model that operates at the paragraph level: inputs are tokenized and padded if they are shorter than the specified input length and truncated if they exceed this limit. To train the model, we merged data from all available languages and performed a split, allocating 75% for training and 25% for validation. We do not apply any other data preprocessing.

For the binary persuasion detection task, we add a binary classification head to the model and fine-tune it to distinguish between persuasive and non-persuasive content. For the persuasion technique classification task (which is a multi-class, multi-label classification problem) we apply a sigmoid

[3] bsnlp.cs.helsinki.fi/shared-task.html

BG		HR		PL		RU		SI	
Team	F_1^m	Team	F_1^m	Team	F_1^m	Team	F_1^m	Team	F_1^m
PSAL_NLP	0.41	Gradient-Flush	0.49	PSAL_NLP	0.42	INSAntive	0.30	Gradient-Flush	0.32
INSAntive	0.34	PSAL_NLP	0.44	Gradient-Flush	0.41	PSAL_NLP	0.29	PSAL_NLP	0.30
Gradient-Flush	0.34	baseline	0.44	INSAntive	0.41	oplot	0.21	baseline	0.27
dutir	0.28	UFAL4DEM	0.36	FactUE	0.39	Gradient-Flush	0.19	INSAntive	0.20
FactUE	0.23	dutir	0.30	dutir	0.36	UFAL4DEM	0.13	dutir	0.19
UFAL4DEM	0.21	INSAntive	0.30	UFAL4DEM	0.25	FactUE	0.11	oplot	0.18
oplot	0.20	oplot	0.27	baseline	0.24	baseline	0.02	UFAL4DEM	0.17
baseline	0.16	FactUE	0.17	oplot	0.20			FactUE	0.08

Table 5: Subtask 2: Micro-averaged F_1 Scores

activation over the output layer to obtain independent class probabilities. Rather than selecting only the highest-scoring label, we adopt a fixed confidence threshold of 0.3, and predict as positive all classes whose probabilities exceed this value. For baselines, we do not apply any hyperparameter tuning. A full list of hyperparameters used to reproduce our baselines is provided in Appendix B. Model selection is based on the F_1 score for the positive class in the binary setting, and the micro-averaged F_1 score in the multi-class, multi-label setting. Our baseline models are publicly available.[4]

6.2 Subtask 1: Detetcion

6.2.1 Results

The official system ranking for subtask 1 is shown in Table 3, with a visual comparison of the systems including the baseline in Figure 3.

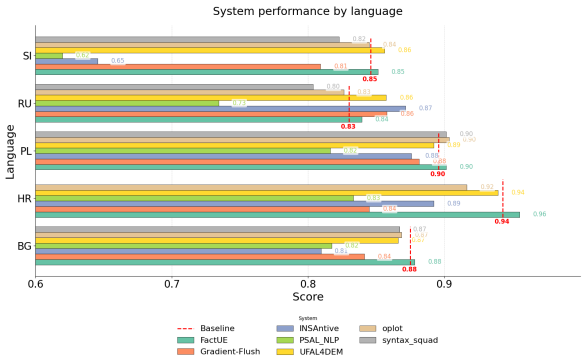

Figure 3: Subtask 1: System performance by language and comparison versus baseline.

6.2.2 System Highlights

Except for Russian, most of the systems perform below the XLM-Roberta baseline. However, most of the systems show to be very close to that baseline, showing this baseline to be the upper bound of the current technology.

FactUE, the best-performing system, probably achieved the upper hand through an auxiliary contrastive learning objective along the main classification task. The team used GPT-4 to generate semantically equivalent, but stylistically neutral text, allowing the model to learn to separate content from persuasion style. With that, the model became less dependent on superficial and stylistic cues, ensuring better generalization to new instances.

Based on the SyntaxSquad (Yahan et al., 2025) submission which used no extra data outperforming others which did, as well as the relatively small margins in performance difference between submissions with varying amounts of extra data, there seems to be no relationship between obtaining additional data and system performance improvements. For this subtask, fine-tuned BERT-like transformer models outperform the single system based purely on LLM prompting, however Loginova (2025) seems to indicate LLM performance on par with their fine/tuned transformer submission.

6.2.3 Comparison of approaches

In Table 6 we observe that most teams used single (one team submitted an ensemble solution) fine-tuned transformer models of the BERT family. Fewer, but still the majority of submissions, used extra training data, focusing mostly on previous similar tasks (Dimitrov et al., 2021; Piskorski et al., 2023b; Dimitrov et al., 2024). Approaches to producing additional training data can be roughly split into two groups: the first one focused on providing more human-made data, the other instead used various techniques to produce synthetic data (either through machine translation or some other generational process). One of the teams submitting transformer-based systems presented an approach using hyperbolic graph convolutional networks. A

[4]Task 1 baseline model: https://huggingface.co/SlavicNLP/SlavRoBERTa-Persuasion-Baseline, Task 2 baseline model: https://huggingface.co/SlavicNLP/SlavRoBERTa-PT-Classification-Baseline

single team submitted LLM-prompt based systems with no training (apart from in-context interventions).

6.3 Subtask 2: Classification

6.3.1 Results

The official system ranking for subtask 2 is shown in Table 4 (macro F_1) and Table 5 (micro F_1), while a visual comparison of the systems in terms of micro and macro F_1, including the baseline, is provided in Figure 5 and 6 respectively. Figure 4 provides a fine-grained comparison of F_1 for each language and team by persuasion technique.

6.3.2 System Highlights

PSAL_NLP: best-performing system, achieved 1st place on both macro F_1 and micro F_1, using chain-of-thought prompts combined with a two-pass technique to split the persuasion techniques into two groups (with a separate prompt for each group). In this way, they address the problem of "cognitive overload" when working with all 25 definitions.

UFAL4DEM: explored hierarchical text classification using graph-based models embedded in hyperbolic space, where the authors model the persuasion label structure from the SemEval-2024 (Dimitrov et al., 2024) task as a graph, with each node representing a technique, and edges reflecting hierarchical relationships.

FactUE: first split the multi-label classification problem into 25 binary classification tasks. Then they introduce a process to refine persuasion technique definitions, which involves GPT-4.1-mini generating “improved” definitions of persuasion techniques, which are then used in the prompt for evaluation. Using this approach, the authors achieve significantly higher results compared to prompting with the original PT definitions.

6.3.3 Comparison of approaches

In Table 7 we observe that the majority of the teams opted for single fine-tuned transformer models, with XLM-RoBERTa being a frequent choice. Almost every team used data from previous shared tasks on persuasion techniques, and 3 teams used machine-translated synthetic data to enrich their dataset. Some teams experimented with automatic data generation, creating explanations of each text sample and combining the newly generated content with the original text to form new training data. Two teams formulate the task as a multi-task problem with 25 binary classifications. Regarding system ranking, we notice differences in macro F_1 and micro F_1 leaderboards, with systems using commercial LLMs dominating the macro F_1 leaderboard, and micro F_1 leaderboards are mostly dominated by single fine-tuned transformer models, namely XLM-RoBERTa. In terms of languages, leaderboard results show that high-ranking systems maintain their performance, except for the Russian leaderboard, where the domain was news articles, indicating that systems are not effective at transferring knowledge from parliamentary debates.

6.4 Discussion

Comparing the results from the current task with recent competitions on persuasion techniques, e.g., (Piskorski et al., 2023b), we can make several general observations. Although we can observe some improvement in the results in terms of the F-measures, these improvements are modest, and results are largely similar the previous years.

For example, for some of the more *frequent* persuasion techniques, we get comparatively better results; they appear to be easier to classify than others. For example: Attacks on Reputation (*Name calling*, *Appeal to hypocrisy*, *Doubt*); Justifications (*Appeal to popularity*, *Appeal to fear or prejudice*); Manipulative Wording (*Exaggeration/minimization*, *Loaded language*)—exhibit better scores than other techniques, both in the current task and those from previous years (Piskorski et al., 2023c). They also have consistently higher support in the dataset. Distractions and Simplifications—which have less support—continue to score low.

Some differences in scores may be due to differences in the sub-corpora; e.g., while scores on *Repetition* are higher than in previous years for Polish and Croatian, in Russian they are very low. This may be because repetitions are not employed as much in social media texts.

We should note that even for the classes that appear “easier” and better supported, performance is still below usable levels. This suggests that classification of persuasion techniques remains a very complex challenge.

Many more of the competing systems rely on LLMs, as compared to earlier competitions. Thus, one way to push progress on this challenge is by increasing the amount of high-quality annotated data. This may allow us to fine-tune LLMs to perform better analysis. However, developing such datasets—and assuring their quality—is very ex-

Reference	Models used	Ensemble	Extra data	Synthetic MT data	Notes
GradientFlush (Senichev et al., 2025)	Slavic-BERT	—	+	+	
UFAL4DEM (Brückner and Pecina, 2025)	XLM-R-parla	—	+	+	Largest training set, uses hyperbolic graph convolutional networks.
INSANTIVE (Wang et al., 2025)	XLM-RoBERTa-base, XLM-RoBERTa-large	—	+	+	Training data augmented with LLM-generated explanations of PTs.
Oplot (Loginova, 2025)	intfloat/multilingual-e5-small	—	+	—	Additional human-generated labels.
FactUE (Księżniak et al., 2025)	jinaai/jina-embeddings-v3, intfloat/multilingual-e5-large	—	+	—	Contrastive loss and model debiasing, training example text pair generated with LLM.
Syntax Squad (Yahan et al., 2025)	BERTić*, XLM-RoBERTa-large, bert-web-bg, herbert, Polbert, Polish-roberta, SloBERTa, SlovakBERT CroSloEngual BERT, Conversational Ru-BERT, RuBERT-tiny, ruBert-base	+	—	—	
PSAL NLP (Jose and Greenstadt, 2025)	gpt-4o-mini and o4-mini with prompt engineering	—	—	—	

Table 6: Comparison of systems for Subtask 1.

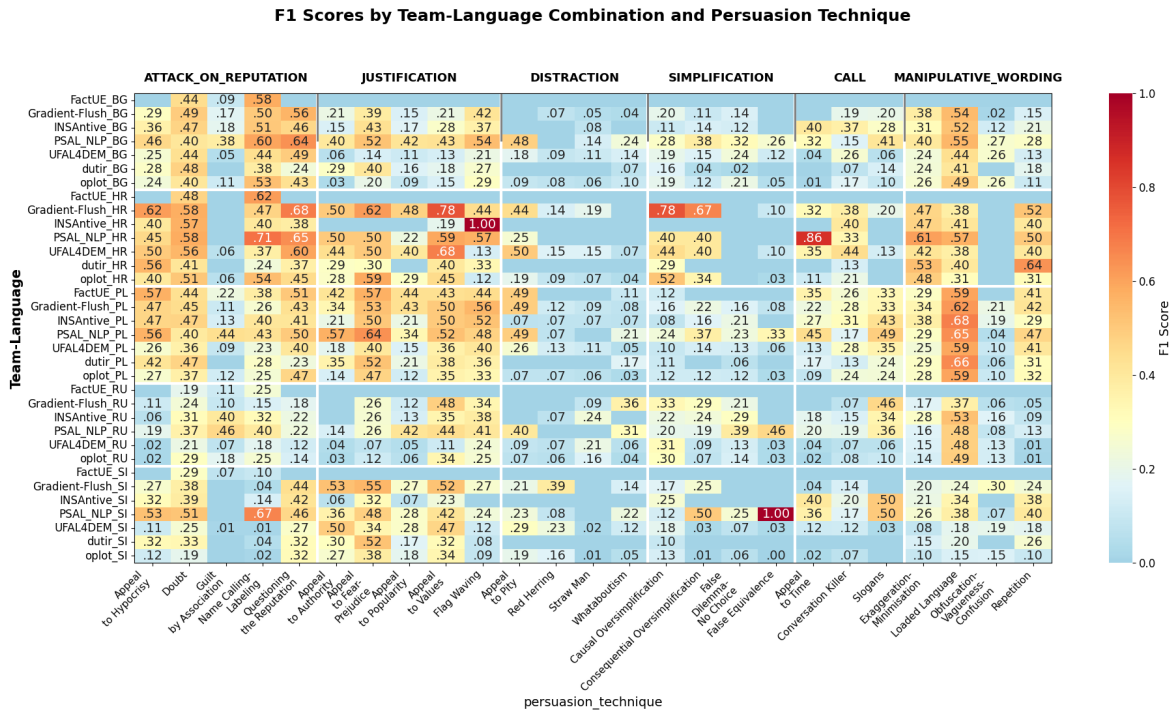

Figure 4: Subtask 2: Fine-grained comparison of F_1 for each language and team by persuasion-technique.

pensive, in terms of human time and effort.

Another avenue may be pattern-based techniques, employed as part of a hybrid approach: language models combined with tools that detect patterns. This is related to research on information extraction, where it is likewise widely recognized that differences in scenarios often correlate with differences in performance (Piskorski and Yangarber, 2013; Huttunen et al., 2002).

Alternatives approaches combining LLMs with "simpler" pattern-based techniques—*hybrid* approaches—are currently an active area of research (Shen et al., 2020; Agarwal et al., 2020). For example, *lexical* features may have strong discriminatory power for some persuasion techniques. Distractions and simplifications are less about lexicon or syntactic patterns. Most systems in previous years are not good at detecting Simplifications, although many examples follow clear syntactic patterns. Thus, combining LLMs with pattern-based

Reference	Models used	Ensemble	Extra data	Synthetic MT data	Notes
DUTIR (Xin et al., 2025)	Qwen3, Qwen2.5 Teacher-student training	+	+	—	Teacher generates explanations, then student learns to approximate them
FactUE (Sawiński et al., 2025)	GPT-4.1-mini, LlaMa3.1, DeepSeek-R1	—	+	—	25 binary classification tasks, PT definition refinement via LLM
GradientFlush (Senichev et al., 2025)	Slavic-BERT, XLM-RoBERTa	—	+	+	
INSANTIVE (Wang et al., 2025)	XLM-RoBERTa, XLM-RoBERTa-large	—	+	+	Training data augmented with LLM-generated explanations of PTs.
Oplot (Loginova, 2025)	XLM-RoBERTa-base	—	+	—	Additional human-generated labels.
UFAL4DEM (Brückner and Pecina, 2025)	XLM-R-parla	—	+	+	Largest training set, uses hyperbolic graph convolutional networks.
PSAL NLP (Jose and Greenstadt, 2025)	gpt-4o-mini, o4-mini, CoT with prompt engineering	—	—	—	25 binary classification tasks

Table 7: Comparison of systems for Subtask 2.

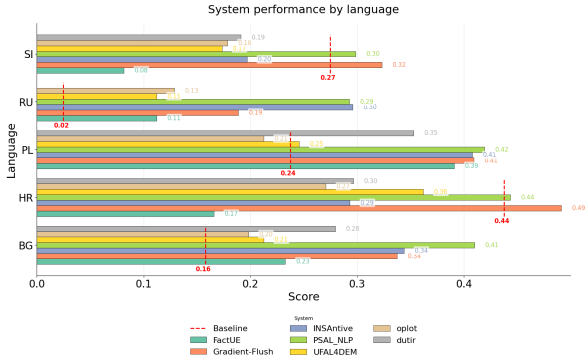

Figure 5: Subtask 2 micro-averaged F_1: Comparison of system performance and baseline.

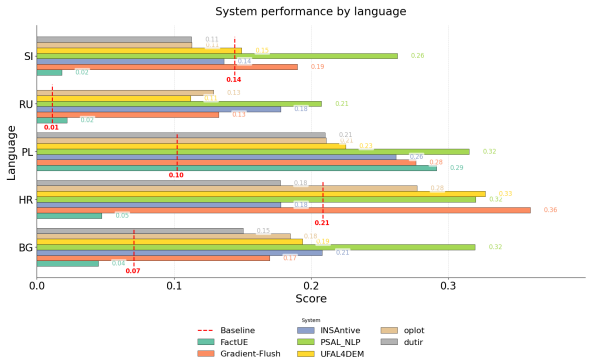

Figure 6: Subtask 2 macro-averaged F_1: Comparison of system performance and baseline.

and knowledge-based techniques may be a fruitful way forward.

7 Conclusions and Future Work

In this paper, we present the SlavicNLP 2025 Shared Task on *Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media*. 15 teams registered to the task, of which 9 teams submitted a total of circa 220 system responses, and described their approaches in 9 system description papers. Although the detection of persuasion techniques at the paragraph level turned out to be a relatively simple task with F_1 scores oscillating around 0.9, the classification of techniques at the paragraph level continues (compared to the previous competition with a similar task formulation) to be a challenging task, where none of the systems achieved a F_1 (micro and macro) score above 0.5.

In the future, we plan to create more annotated data and include other Slavic languages, and exploit the existing data to explore solutions for other related tasks, e.g., detection and classification of persuasion techniques at the sentence level, and detection of political bias.

8 Ethics Policy

Intended Use and Misuse Potential: The data sets created in the context of the presented Shared Task were designed to advance research on detection and classification of persuasion techniques for the domain of parliamentary debates and social media. Given the potential risks of exploiting these data sets for the production of manipulative content, we strongly advise responsible use of the data.

Fairness: We engaged a number of annotators to create the data sets for this Shared Task. Some are researchers with a (computational) linguistic and/or social sciences background and prior annotation experience, coming from the institutions of the co-organizers of the Task. They were fairly remunerated as part of their job.

Other annotators were (a) students from the re-

spective academic organizations, and (b) experts from a contracted professional annotation company, who were compensated according to rates based on their country of residence.

9 Limitations

Dataset Representativeness: The datasets used in our shared task cover parliamentary debates and propaganda narratives in various countries and we strove to include utterances of speakers covering a wide political spectrum in each of these countries. However, we must emphasize that these datasets should not be considered representative of the political landscape in any specific country or region, nor should they be considered as balanced in any way.

Biases: We have invested a significant effort in training the annotators and acquainting them with the specifics of the persuasion technique taxonomy. Furthermore, cross-language quality control mechanisms have been put in place to ensure the highest quality of annotations. Nevertheless, some degree of intrinsic subjectivity might be present in the datasets. Therefore, models trained using these datasets might exhibit certain biases.

Acknowledgements

We express deep gratitude to DataBee (get-databee.com/) and specifically Peter-Michael Slaveykov, Krasen Zhelyzkov, Samuil Ivanov, and Blagovest Chernev for their invaluable contribution to the annotation of Bulgarian data. We are thankful to the Croatian data annotators: Karlo Kralj, Mirna Potočnjak, and Maja Živković.

We are very grateful to the University of Helsinki team for annotation of the Russian data: Denis Kvachev, Irina Gatsuk and Matilda Villanen. This work was in part supported by the Research Council of Finland.

This research is partially funded by the EU NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project SUMMIT, No BG-RRP-2.004-0008. Partial funding was also obtained through the Research Programme "Digital humanities: resources, tools and methods" (P6-0436), and the Research Infrastructure DARIAH-SI (I0-E007), both funded by the Slovenian Research and Innovation Agency ARIS.

References

Gavin Abercrombie and Riza Batista-Navarro. 2020. Sentiment and position-taking analysis of parliamentary debates: a systematic literature review. *Journal of Computational Social Science*, 3(1):245–270.

Oshin Agarwal, Heming Ge, Siamak Shakeri, and Rami Al-Rfou. 2020. Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pre-training. *arXiv preprint arXiv:2010.12688*.

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 108–118, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Christopher Brückner and Pavel Pecina. 2025. Hierarchical classification of propaganda techniques in Slavic texts in hyperbolic space. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

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.

Giovanni Da San Martino, Alberto Barrón-Cedeño, and Preslav Nakov. 2019. Findings of the NLP4IF-2019 shared task on fine-grained propaganda detection. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 162–170, Hong Kong, China. 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.

Dimitar Dimitrov, Firoj Alam, Maram Hasanain, Abul Hasnat, Fabrizio Silvestri, Preslav Nakov, and Giovanni Da San Martino. 2024. SemEval-2024 task 4: Multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 2009–2026, Mexico City, Mexico. 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.

Tomaž Erjavec, Matyáš Kopp, Nikola Ljubešić, Taja Kuzman, Paul Rayson, Petya Osenova, Maciej Ogrodniczuk, Çağrı Çöltekin, Danijel Koržinek, Katja Meden, et al. 2024. ParlaMint II: advancing comparable parliamentary corpora across Europe. *Language Resources and Evaluation*, pages 1–32.

Bojan Evkoski and Senja Pollak. 2023. XAI in computational linguistics: Understanding political leanings in the Slovenian parliament. *arXiv preprint arXiv:2305.04631*.

Maram Hasanain, Firoj Alam, Hamdy Mubarak, Samir Abdaljalil, Wajdi Zaghouani, Preslav Nakov, Giovanni Da San Martino, and Abed Freihat. 2023. ArAIEval shared task: Persuasion techniques and disinformation detection in Arabic text. In *Proceedings of ArabicNLP 2023*, pages 483–493, Singapore (Hybrid). Association for Computational Linguistics.

Maram Hasanain, Md. Arid Hasan, Fatema Ahmad, Reem Suwaileh, Md. Rafiul Biswas, Wajdi Zaghouani, and Firoj Alam. 2024. ArAIEval shared task: Propagandistic techniques detection in unimodal and multimodal Arabic content. In *Proceedings of the Second Arabic Natural Language Processing Conference*, pages 456–466, Bangkok, Thailand. Association for Computational Linguistics.

Silja Huttunen, Roman Yangarber, and Ralph Grishman. 2002. Diversity of scenarios in information extraction. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Spain.

Julia Jose and Rachel Greenstadt. 2025. LLMs for detection and classification of persuasion techniques in Slavic parliamentary debates and social media texts. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics. Event Title: The 27th International Conference on Computational Linguistics (COLING 2018).

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of machine translation summit x: papers*, pages 79–86.

Ewelina Księżniak, Krzysztof Węcel, and Marcin Sawiński. 2025. Robust detection of persuasion techniques in Slavic languages via multitask debiasing and walking embeddings. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Nikola Ljubešić and Peter Rupnik. 2002. The parlaspeech-hr benchmark for speaker profiling in croatian.

Nikola Ljubešić, Peter Rupnik, and Danijel Koržinek. 2024. The parlaspeech collection of automatically generated speech and text datasets from parliamentary proceedings. In *International Conference on Speech and Computer*, pages 137–150. Springer.

Ekaterina Loginova. 2025. Fine-tuned transformers for detection and classification of persuasion techniques in Slavic languages. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Michal Mochtak, Peter Rupnik, and Nikola Ljubešić. 2024. The ParlaSent Multilingual Training Dataset for Sentiment Identification in Parliamentary Proceedings. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16024–16036.

Pablo Moral, Jesús M Fraile, Guillermo Marco, Anselmo Peñas, and Julio Gonzalo. 2024. Overview of dipromats 2024: Detection, characterization and tracking of propaganda in messages from diplomats and authorities of world powers. *Procesamiento del lenguaje natural*, 73:347–358.

Pablo Moral, Guillermo Marco, Julio Gonzalo, Jorge Carrillo-de Albornoz, and Iván Gonzalo-Verdugo. 2023. Overview of dipromats 2023: automatic detection and characterization of propaganda techniques in messages from diplomats and authorities of world powers. *Procesamiento del lenguaje natural*, 71:397–407.

Jakub Piskorski, Nicolas Stefanovitch, Firoj Alam, Ricardo Campos, Dimitar Dimitrov, Alípio Jorge, Senja Pollak, Nikolay Ribin, Zoran Fijavz, Maram Hasanain, Purificação Silvano, Elisa Sartori, Nuno Guimarães, Ana Zwitter Vitez, Ana Filipa Pacheco, Ivan Koychev, Nana Yu, Preslav Nakov, and Giovanni Da San Martino. 2024. Overview of the CLEF-2024 checkthat! lab task 3 on persuasion techniques. In *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024*, volume 3740 of *CEUR Workshop Proceedings*, pages 299–310. CEUR-WS.org.

Jakub Piskorski, Nicolas Stefanovitch, Valerie-Anne Bausier, Nicolo Faggiani, Jens Linge, Sopho Kharazi, Nikolaos Nikolaidis, Giulia Teodori, Bertrand De Longueville, Brian Doherty, Jason Gonin, Camelia Ignat, Bonka Kotseva, Eleonora Mantica, Lorena Marcaletti, Enrico Rossi, Alessio Spadaro, Marco Verile, Giovanni Da San Martino, Firoj Alam,

and Preslav Nakov. 2023a. News categorization, framing and persuasion techniques: Annotation guidelines. Technical report, European Commission Joint Research Centre, Ispra (Italy).

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023b. 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)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

Jakub Piskorski, Nicolas Stefanovitch, Nikolaos Nikolaidis, Giovanni Da San Martino, and Preslav Nakov. 2023c. Multilingual multifaceted understanding of online news in terms of genre, framing, and persuasion techniques. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3001–3022, Toronto, Canada. Association for Computational Linguistics.

Jakub Piskorski and Roman Yangarber. 2013. Information extraction: past, present and future. In *Multi-source, multilingual information extraction and summarization*, pages 23–49. Springer.

Marcin Sawiński, Krzysztof Węcel, and Ewelina Księżniak. 2025. Multilabel classification of persuasion techniques with self-improving LLM agent: SlavicNLP 2025 Shared Task. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Miklós Sebők, Ákos Máté, Orsolya Ring, Viktor Kovács, and Richárd Lehoczki. 2024. Leveraging open large language models for multilingual policy topic classification: The Babel machine approach. *Social Science Computer Review*.

Sergey Senichev, Aleksandr Boriskin, Nikita Krayko, and Daria Galimzianova. 2025. Gradient Flush at Slavic NLP 2025 Task: Leveraging Slavic BERT and translation for persuasion techniques classification. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Tao Shen, Yi Mao, Pengcheng He, Guodong Long, Adam Trischler, and Weizhu Chen. 2020. Exploiting structured knowledge in text via graph-guided representation learning. *arXiv preprint arXiv:2004.14224*.

Nicolas Stefanovitch and Jakub Piskorski. 2023. Holistic inter-annotator agreement and corpus coherence estimation in a large-scale multilingual annotation campaign. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 71–86, Singapore. Association for Computational Linguistics.

Yutong Wang, Diana Nurbakova, and Sylvie Calabretto. 2025. Team INSAntive at SlavicNLP-2025 Shared Task: Data augmentation and enhancement via explanations for persuasion technique classification. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Zou Xin, Wang Chuhan, Li Dailin, Wang Yanan, Wang Jian, and Lin Hongfei. 2025. Empowering persuasion detection in Slavic texts through two-stage generative reasoning. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Mahshar Yahan, Sakib Sarker, and Mohammad Amanul Islam. 2025. Fine-tuned transformer-based weighted ensemble for binary classification in Slavic languages. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

A Persuasion Techniques Definitions and Examples

Below we provide the definitions of the persuasion techniques accompanied by examples in English (in blue) and in the Slavic languages (in brown) of the Shared Task. The text fragments highlighted in bold are the text spans to be annotated according to the guidelines presented in (Piskorski et al., 2023a).

The definitions of the persuasion technique are taken directly from the Annex of (Piskorski et al., 2023c), with two new persuasion techniques: *Appeal to Pity* and *False Equivalence*, which were added for this task.

A.1 Attack on Reputation

Name Calling or Labeling: a form of argument in which loaded labels are directed at an individual or a group, typically in an insulting or demeaning way. An object is labeled as something the target audience fears, hates, or, on the contrary, finds desirable or loves. This technique calls for a qualitative judgement that disregards facts and focuses solely on the essence of the subject being characterized. This technique is also in a way manipulative wording, as it appears as a nominal group rather than being a full-fledged argument with a premise and a conclusion. For example, in political discourse, typically one uses adjectives and nouns as labels that refer to political orientation, opinions, personal characteristics, and association to some organisations, as well as insults. What distinguishes it from

Loaded Language (see A.6), is that it is concerned only with the characterization of the subject.
Example: ***'Fascist' Anti-Vax** Riot Sparks COVID Outbreak in Australia.*
Example: *Trzeba zrozumieć, że bronią także i polskich granic przeciwko **rosyjskiemu imperializmowi**, którego ducha wskrzesił Władimir Putin—**prezydent zbrodniarz**.* (*It is necessary to understand that they are also defending the Polish borders against **Russian imperialism**, whose spirit has been revived by Vladimir Putin—**the criminal president.***)

Guilt by Association: Attacking an opponent or an activity by associating it with another group, activity, or concept that has sharply negative connotations for the target audience. The most common example, which has given its name in the literature to this technique (i.e., *Reduction ad Hitlerum*) is making comparisons with Hitler and the Nazi regime. However, it is important to emphasize, that this technique is not restricted to comparisons to that group only. More precisely, this can be done by claiming a link or an equivalence between the target of the technique and any individual, group, or event in the present or in the past, which is or was negatively perceived (e.g., was considered a failure), or is depicted in such a way.
Example: ***Manohar is a big supporter for equal pay for equal work. This is the same policy that all those extreme feminist groups support.** Extremists like Manohar should not be taken seriously.*
Example: *Мы часто забываем, что после Второй мировой наши типа союзники, французы **(на самом деле настоящие союзники Гитлера)**, стали срочно восстанавливать свою империю.* (*We often forget that after WWII our so-called allies, the French **(Hitler's allies, actually)**, immediately started rebuilding their empire.*)

Casting Doubt: Casting doubt on the character or the personal attributes of someone or something in order to question their general credibility or quality, rather than using a proper argument relevant to the topic. This can be done for instance, by speaking about the target's professional background, as a way to discredit their argument. Casting doubt can also be done by referring to some actions or events carried out or planned by some entity that are/were not successful, or appear as resulting in not achieving the planned goals.
Example: *This task is quite complex. **Is his professional background, experience and the time left sufficient to accomplish the task at hand?***
Example: ***Predlagatelji v očitkih o delu NPU ne govorijo ne o dejstvih in ne o dokazih.*** (***In their accusations regarding the work of NPU the proponents speak neither of facts nor evidence.***)

Appeal to Hypocrisy: The reputation of the target is attacked by charging them with hypocrisy or inconsistency. This can be done explicitly by calling out hypocrisy directly, or implicitly by underlining the contradictions between different positions that were held or actions that were done in the past. A common way of calling out hypocrisy is by saying that someone who criticizes you for something you have done, has done it himself in the past.
Example: ***How can you demand that I eat less meat to reduce my carbon footprint if you yourself drive a big SUV and fly for holidays to Bali?***
Example: *Иначе ЦЕМ твърди, че е безпристрастен, но когато става въпрос за безпочвени обвинения към Русия или манипулиране на общественото мнение по този начин, някак си ЦЕМ пропуска това.* (***Otherwise, the CEM claims to be impartial, but when it comes to groundless accusations against Russia or manipulating public opinion in this way, the CEM somehow misses the mark.***)

Questioning the Reputation: This technique is used to attack the reputation of the target by making strong negative claims about it, focusing on undermining its character and moral stature rather than relying on an argument about the topic. Whether the claims are true is irrelevant for the effective use of this technique. Smears can be used at any point in a discussion. One way of using this technique is to preemptively call into question the reputation/-credibility of an opponent, before he has a chance to express himself, therefore biasing the audience's perception. Hence, one of the names for this technique is "poisoning the well."

The main difference between *Casting Doubt* (above) and *Questioning the reputation* is that the former focuses on questioning the capacity, capabilities, and credibility of the target, while the latter aims to undermine the overall reputation, moral qualities, behaviour, etc.
Example: *I hope I presented my argument clearly. Now, **my opponent will attempt to refute my argument by his own fallacious, incoherent, illogical version of history.***

Example: *A ta ministrica je lagala, lagala, lagala matičnom saborskom odboru kada je odgovarala na pitanja tko je zakon pisao i sastavljao.* (***But the minister lied, lied, lied to the working body when she was answering questions about who wrote the law and put it together.***)

A.2 Justification

Flag Waving: Justifying or promoting an idea by appealing to the pride of a group or highlighting the benefits for that specific group. The stereotypical example would be national pride, and hence the name of the technique; however, the target may be any group, e.g., related to race, gender, political preference, etc. The connection to nationalism, patriotism, or benefit for an idea, group, or country might be inappropriate and is usually based on the presumption that the recipients already hold certain beliefs, biases, and prejudices about the given issue. It can be seen as an appeal to emotions instead to logic of the audience aiming to manipulate them to win an argument. As such, this technique can also appear outside well-constructed arguments, by making statements that resonate with the particular group and as such setting up a context for further arguments.
Example: ***We should make America great again, and restrict the immigration laws.***
Example: *Wolna Ukraina i silna Unia Europejska, silna Polska stanowią podstawę polskiej racji stanu, to podstawa naszego bezpieczeństwa.* (***A free Ukraine and a strong European Union, a strong Poland, are the foundation of the Polish national interest, they are the basis of our security.***)

Appeal to Authority: attempting to add weight to an argument, an idea or information by simply stating that a particular entity considered to be an authority is the source of the information. The entity mentioned as an authority may, but does not need to be, an actual authority in the specific domain to discuss a particular topic or to serve as an expert. What is important, and makes it different from simply sourcing information, is that the tone of the text capitalizes on the weight of the alleged authority in order to justify some claim or conclusion. Referencing a valid authority is not a logical fallacy, while referencing an invalid authority is a logical fallacy, and both are captured within this label. In particular, a self-reference as an authority falls under this technique as well.
Example: ***Since the Pope said that this aspect of the doctrine is true we should add it to the creed.***
Example: *Strokovnjaki dnevno opozarjajo, da se je duševno zdravje posameznikov med pandemijo poslabšalo, duševne stiske pa se bodo povečevale še dolgo po njenem koncu.* (***Every day we hear warnings from experts saying the mental health of individuals has deteriorated during the pandemic, and mental distress will continue to intensify long after the pandemic is over.***)
Example: *Глава ЦБ РФ Эльвира Набиуллина назвала новые реалии тектоническими изменениями в мировой торговле, и с учётом всех нюансов происходящего это ещё очень деликатная формулировка.* (***The head of the Central Bank of Russia Elvira Nabiullina called the new situation a "tectonic shift in global trade," and considering all the nuances of what is happening, this is still a very delicate formulation.***)

Appeal to Popularity: This technique gives weight to an argument or idea by justifying it on the basis that allegedly "*everyone*" (or the vast majority) agrees with it, or "*nobody*" disagrees with it. The target audience is encouraged to gregariously adopt the same idea by considering "*everyone*" as an authority, and to join in and take the same course of action. Here, "*everyone*" might refer to the general public, key entities and actors in a certain domain, countries, etc. Analogously, an attempt to persuade the audience not to do something because "*nobody else is taking the same action*" falls under our definition of *Appeal to Popularity*.
Example: ***Because everyone else goes away to college, it must be the right thing to do.***
Example: *Stroka, mediji, novinarji, politiki so rekli, da je to odlično, morda celo najbolje pripravljena interpelacija do sedaj.* (***Experts, media, journalists, politicians all said that this is outstanding, maybe even the best prepared interpellation until now.***)

Appeal to Values: This technique gives weight to an idea by linking it to values seen by the target audience as positive. These values are presented as an authoritative reference in order to support or to reject an argument. Examples of such values are, for instance: tradition, religion, ethics, age, fairness, liberty, democracy, peace, transparency, etc. When such values are mentioned outside the context of a proper argument by simply using cer-

tain adjectives or nouns as a way of characterizing something or someone, such references fall under another label, namely, *Loaded Language*, which is a form of *Manipulative Wording* (see A.6).
Example: ***It's standard practice to pay men more than women so we'll continue adhering to the same standards this company has always followed.***
Example: *В очередной раз удар нанесён по одной из самых чувствительных сфер—религиозным правам и свободам.* (***Another attack has been made on one of the most sensitive areas—religious rights and freedoms.***)

Appeal to Fear, Prejudice: This technique aims at promoting or rejecting an idea through the repulsion or fear the audience feels toward this idea (e.g., via exploiting some preconceived judgements) or toward its alternative. The alternative could be the status quo, in which case the current situation is described in a scary way with *Loaded Language*. If the fear is linked to the consequences of a decision, it is often the case that this technique is used simultaneously with *Appeal to Consequences* (see Simplification techniques in A.4), and if there are only two alternatives that are stated explicitly, then it is used simultaneously with the *False Dilemma* technique (see A.4).
Example: *It is a great disservice to the Church to maintain the pretense that there is nothing problematical about Amoris laetitia.* ***A moral catastrophe is self-evidently underway*** *and it is not possible honestly to deny its cause.*
Example: *Много, много други такива неща са се случвали и за съжаление,* ***ние отиваме по едни стъпки, които са изключително опасни, изключително наистина тревожни за бъдещето на нашата държава.*** (*Many, many other such things have happened, and unfortunately,****we are taking extremely dangerous steps, extremely worrying for the future of our country.***)

A.3 Distraction

Strawman: This technique consists in creating an illusion of refuting the argument of the opponent's proposition, while the real subject of the argument was not addressed or refuted, but instead replaced with a false one. Often, this technique is referred to as a misrepresentation of the argument. First, a new argument is created via the covert replacement of the original argument with something that appears related, but is actually a different, distorted, exaggerated, or misrepresented version of the original proposition, which is referred to as "*setting up a strawman*." Subsequently, the newly created '*false* argument (strawman) is refuted, which is referred to as "*knocking down the strawman*." Often, the strawman argument is created in such a way that it is easier to refute, and thus, creating the illusion of having defeated an opponent's real proposition. Fighting a strawman is easier than fighting a real person, which explains the name of this technique. In practice, it appears often as an abusive reformulation or explanation of what the opponent *actually* means or intends.
Example: *Referring to your claim that providing medicare for all citizens would be costly and a danger to the free market, I infer* ***that you don't care if people die from not having healthcare, so we are not going to support your endeavour****.*
Example: *Има огромно значение, господин Иванов, дали българското знаме е отляво, или отдясно. Това нещо го знаете по протокол.* ***Ако казвате, че няма значение, това означава, че за Вас няма значение какъв точно ще бъде статутът на българското знаме в България, статутът на българския държавен герб и къде точно ще се полага*** (*It makes a huge difference, Mr Ivanov, whether the Bulgarian flag is on the left or the right. You know this from protocol.****If you say that it does not matter, it means that it does not matter to you exactly what the status of the Bulgarian flag will be in Bulgaria, the status of the Bulgarian state coat of arms and exactly where it will be placed.***)

Red Herring: This technique consists in diverting the attention of the audience from the main topic being discussed, by introducing another topic. The aim of attempting to redirect the argument to another issue is to focus on something the person doing the redirecting can better respond to or to leave the original topic unaddressed. The name of that technique comes from the idea that a fish with a strong smell (such as a herring) can be used to divert dogs from the scent of someone they are following. A strawman (defined earlier) is a specific type of a red herring in that it distracts from the main issue by presenting the opponent's argument in an inaccurate light.
Example: *Lately, there has been a lot of criticism regarding the quality of our product.* ***We've decided***

to have a new sale in response, so you can buy more at a lower cost!.
Example: *Недавно она прочитала лекцию о необходимости войны с ухоженным газоном, потому что «это символ сексизма, расизма и экологического разрушения». **Среди друзей Аджубей много проукраинских активистов и адептов движений Black lives matter и ЛГБТ.*** (*She recently gave a lecture on the need for a war on manicured lawns because "they are a symbol of sexism, racism and ecological destruction" **Adzhubey's friends include many pro-Ukrainian activists and adherents of the Black lives matter and LGBT movements.***)

Whataboutism: Attempt to discredit an opponent's position by charging them with hypocrisy without directly disproving their argument. Rather than answering a critical question or argument, an attempt is made to retort with a critical counter-question that expresses a counter-accusation, e.g., mentioning double standards, etc. The intent is to distract from the content of a topic and to actually switch the topic. There is a fine distinction between this technique and *Appeal to Hypocrisy*, introduced earlier: the former is an attack on the argument and introduces irrelevant information to the main topic, while the latter is an attack on reputation and highlights the hypocrisy of double standards on the same or a closely related topic.
Example: ***A nation deflects criticism of its recent human rights violations by pointing to the history of slavery in the United States.***
Example: *Добре, на Хърватия е пораснал—окей. **А Естония и Финландия, които са на минус, и Ирландия, които са в еврозоната, какво правим?*** (*Okay, Croatia's has grown—okay. **And what about Estonia and Finland, which are in the red, and Ireland, which are in the eurozone, what do we do?***)

Appeal to Pity: Evokes feelings of pity, sympathy, compassion or guilt in audience to distract it from focusing on evidence, rational analysis and logical reasoning, so that it accepts the speaker's conclusion as truthful solely based on the aforementioned emotions. It is an attempt to sway opinions and fully substitute logical evidence in an argument with a claim intended to elicit pity or guilt.
Example: ***If this person is found guilty of this crime, his ten children will be left without a parent at home, therefore the jury must submit a verdict of innocence.***
Example: *Напуганные, изнурённые отсутствием спокойствия и элементарных условий для жизни, женщины всё равно не были сломлены и не потеряли надежду на освобождение российскими подразделениями их родного хутора.* (***Frightened, exhausted by the insecurity and lack of basic living conditions**, the women were still not broken and did not lose hope for the liberation of their village by Russian troops*)

A.4 Simplification

Causal Oversimplification: Assuming a single cause or reason when there are actually multiple causes for an issue. This technique has the following logical form(s): (a) *Y occurred after X; therefore, X was the only cause of Y*, or (b) *X caused Y; therefore, X was the only cause of Y (although A, B, C...etc. also contributed to Y.)*
Example: *School violence has gone up and academic performance has gone down since video games featuring violence were introduced. **Therefore, video games with violence should be banned, resulting in school improvement.***
Example: ***Građani moraju znati lockdown je prvi razlog i euro sad drugi razlog ovakvih cijena.*** (***Citizens must know that the lockdown is the first and the Euro the second reason for these prices.***)

False Dilemma or No Choice: Sometimes called the *either-or* fallacy, a false dilemma is a logical fallacy that presents only two options or sides when there are actually many. One of the alternatives is depicted as a *no-go* option, hence the only choice is the other option. In extreme cases, the author tells the audience exactly what actions to take, eliminating any other possible choices (also referred to as *Dictatorship*).
Example: ***There is no alternative to Pfizer Covid-19 vaccine. Either one takes it or one dies.***
Example: ***Bodisi se upokojijo ali pa si poiščejo boljšo zaposlitev in podajo odpoved.*** (***They either retire or find a better job and quit.***)

Consequential Oversimplification: An argument or an idea is rejected and instead of discussing whether it makes sense and/or is valid, the argument affirms, without proof, that accepting the proposition would imply accepting other propositions that are considered negative. This technique has the following logical form: *if A will happen then B, C, D, ... will happen*. The core essence

behind this fallacy is an assertion one is making of some '*first*' event/action leading to a domino-like chain of events that have some significant negative effects and consequences that appear to be ludicrous. This technique is characterized by *ignoring and/or understating the likelihood of the sequence of events from the first event leading to the end point* (last event). In order to take into account symmetric cases, i.e., using *Consequential Oversimplification* to promote or to support certain action in a similar way, we also consider cases when the sequence of events leads to positive outcomes (i.e., encouraging people to undertake a certain course of action(s), with the promise of a major positive event in the end).

Example: ***If we begin to restrict freedom of speech, this will encourage the government to infringe upon other fundamental rights, and eventually this will result in a totalitarian state where citizens have little to no control of their lives and decisions they make.***

Example: *Сокрытие правды и подмена понятий приведет к тому, что управлять умами и историей будет противник на нашей территории, выдавая правду с нужным ему уклоном.* (***Concealing the truth and substituting concepts will result in the enemy controlling minds and history on our territory, spreading the truth with an intended bias.***)

False Equivalence: A technique that attempts to treat scenarios that are significantly different as if they had equal merit or significance. In particular, an emphasis is placed on one specific shared characteristic between the items of comparison in the argument that is off by an order of magnitude, oversimplified, or important additional factors have been ignored. The introduction of certain shared characteristics of the scenarios is then used to consider them equivalent. This technique has the following logical form: *A and B share some characteristic X. Therefore, A and B are equivalent.*

Example: ***The introduction or restrictive hours of alcohol sales boosted the black market industry, and analogously, one can expect that the introduction of too restrictive anti-abortion regulations will lead to growth of the illegal abortion business.***

Example: ***To właśnie Führer jako pierwszy wprowadził wolną aborcję dla Polek oraz dla innych kobiet z narodów podbitych. Chodziło o fizyczne zniszczenie ludności niearyjskiej i zdobycie lebensraumu dla Niemców. Hitler rozumiał, że jeśli zalegalizuje aborcję, stanie się ona zjawiskiem masowym i spowoduje spadek urodzeń. Na ziemiach podbitych przez Niemcy dzieci niearyjskie uważano za zagrożenie, więc wdrażano politykę sprzyjającą aborcji. Równocześnie za to samo, za zabicie dziecka niemieckiego w Niemczech groziła kara śmierci. A dyktator groził: osobiście zastrzelę tego idiotę, który chciałby wprowadzić w życie przepisy zabraniające aborcji na wschodnich terenach okupowanych. Jaka jest analogia? Kto powiedział: każda odmowa aborcji będzie zgłaszana do prokuratury? Premier rządu rewolucji*** (***It was the Führer who first introduced free abortion for Polish women and other women from conquered nations. The idea was to physically destroy the non-Aryan population and gain Lebensraum for the Germans. Hitler understood that if he legalized abortion, it would become a mass phenomenon and cause a decrease in births. In the lands conquered by Germany, non-Aryan children were considered a threat, so a policy favoring abortion was implemented. At the same time, the same thing, killing a German child in Germany, was punishable by death. And the dictator threatened: I will personally shoot this idiot who would want to implement regulations prohibiting abortion in the occupied eastern territories. What is the analogy? Who said: every refusal to have an abortion will be reported to the prosecutor's office? The prime minister of the government of the revolution***)

Example: *В 1990-х годах были скинхеды—группы асоциальной молодежи, которые толпой нападали на лиц неевропейской наружности, на так сказать «черных».* ***Теперь скинхеды—это группы асоциальной молодежи среднеазиатской наружности, которые так и не смогли гармонично жить рядом с русскими, и толпой избивают русских парней и насилуют русских девочек*** (*In the 1990s, there were the skinheads—groups of antisocial youth who mobbed people of non-European appearance, the so-called "blacks".* ***Now skinheads are groups of antisocial youth of Central Asian appearance, who failed to live peacefully next to Russians and beating Russian guys and raping Russian girls***)

A.5 Call

Slogans: A brief and striking phrase that may include labeling and stereotyping. Slogans tend to act as emotional appeals.
Example: ***Immigrants welcome, racist not!***
Example: *Да живее България!* (***Long live Bulgaria!***)

Conversation Killer: This includes words or phrases that discourage critical thought and meaningful discussion about a given topic. They are a form of *Loaded Language*, often passing as folk wisdom, intended to end an argument and quell cognitive dissonance.
Example: *I'm not so naïve or simplistic to believe we can eliminate wars.* ***You can't change human nature.***
Example: ***Takie są fakty i taka jest polska racja stanu.*** (***TThese are the facts, and this is the Polish national interest.***)

Appeal to Time: The argument is centered around the idea that the time has come for a particular action. The very timeliness of the idea is part of the argument.
Example: *This is no time to engage in the luxury of cooling off or to take the tranquilizing drug of gradualism.* ***Now is the time to make real the promises of democracy. Now is the time to rise from the dark and desolate valley of segregation to the sunlit path of racial justice.***
Example: ***A krajnje je vrijeme da se to ukine ili barem preispita.*** (***It is high time for this to be shut down or at least questioned.***)

A.6 Manipulative Wording

Loaded Language: use of specific words and phrases with strong emotional implications (either positive or negative) to influence and to convince the audience that an argument is valid. It is also known as *Appeal to Argument from Emotive Language*.
Example: *They keep feeding these people with* ***trash****. They should stop.*
Example: ***Nękanie*** *zasłużonej dla szerzenia polskości instytucji* ***bezzasadnymi*** *pozwami odbierane jest m.in. przez moich wyborców jako działania mające na celu* ***sparaliżowanie*** *funkcjonowania tej fundacji.* (*The* ***harassment*** *of an institution that has earned merit in promoting Polish identity through* ***groundless*** *lawsuits is perceived, among others by my constituents, as actions aimed at* ***paralyzing*** *the functioning of this foundation.*)

Obfuscation, Intentional Vagueness, Confusion: This fallacy uses words that are deliberately unclear, so that the audience may have its own interpretations. For example, an unclear phrase with multiple or unclear definitions is used within the argument and, therefore, does not support the conclusion. Statements that are imprecise and intentionally do not fully or vaguely answer the posed question fall under this category.
Example: ***Feathers cannot be dark, because all feathers are light!***
Example: ***Izvajamo ukrepe za pospešeno pridobivanje in zadrževanje kadrov ter razvijamo inovativne pristope zaposlovanja, podprte z informacijskimi tehnologijami.*** (***We are implementing measures for sped up reception and retention of human resources and developing innovative approaches to hiring, supported with information technology.***)

Exaggeration or Minimisation: This technique consists of either representing something in an excessive manner—by making things larger, better, worse (e.g., *the best of the best*, *quality guaranteed*)—or by making something seem less important or smaller than it really is (e.g., saying that an insult was just a joke), downplaying the statements and ignoring the arguments and the accusations made by an opponent.
Example: *From the seminaries, to the clergy, to the bishops, to the cardinals,* ***homosexuals are present at all levels, by the thousand****.*
Example: ***Europa prowadzi również najbardziej dramatyczną wojnę, wojnę demograficzną, którą przegrywa.*** (***Europe is also fighting its most dramatic war, the demographic war, which it is losing.***)

Repetition: The speaker uses the same word, phrase, story, or imagery repeatedly in the hope that the repetition will persuade the audience.
Example: ***Hurtlocker deserves an Oscar****. Other films have potential, but they do not* ***deserve an Oscar like Hurtlocker does****. The other movies may deserve an honorable mention but* ***Hurtlocker deserves the Oscar****.*
Example: ***Da li mi stvarno želimo*** *imati konsolidaciju?* ***Da li želimo stvarno*** *da se ta konsolidacija provede?* ***Da li zaista želimo*** *riješiti taj problem?* (***Do we really want*** *to have consolidation?* ***Do we really want*** *for this consolidation to go through?* ***Do we really want*** *to solve this problem?*)

Hyperparameter	Value
Max input length	128
Batch size (train/eval)	16 / 16
Number of epochs	3
Learning rate	5e-5
Evaluation steps	100
Mixed precision (FP16)	True

Table 8: Hyperparameters used for fine-tuning the *XLM-RoBERTa-base* model on Task 1 (binary persuasion detection).

Hyperparameter	Value
Max input length	256
Batch size (train/eval)	8 / 8
Number of epochs	8
Learning rate	5e-5
Evaluation steps	50
Mixed precision (FP16)	True

Table 9: Hyperparameters used for fine-tuning the *XLM-RoBERTa-base* model on Task 2 (multi-label multi-class persuasion classification).

B Details of Baseline Systems

For both Task 1 and Task 2, we fine-tuned the multilingual *XLM-RoBERTa-base* model using the official datasets provided as part of the SlavicNLP 2025 shared task. Each model was trained with a task-specific set of hyperparameters. Table 8 outlines the hyperparameters used for binary persuasion detection (Task 1), while Table 9 lists those used for multi-label, multi-class persuasion technique classification (Task 2). The resulting models are publicly available on the Hugging Face Hub:

- *SlavRoBERTa-Persuasion-Baseline*
- *SlavRoBERTa-PT-Classification-Baseline*

C Participant Systems

In this section, we list all participants who submitted a system description. The team name used for the submission is in bold. The list of subtasks the team participated in is given in brackets. A short description of the system is provided.

DUTIR [ST2] (Xin et al., 2025) (Keywords: *Qwen3, Qwen2.5, Teacher-student training, Ensemble, Fine-tuning, Automatic data generation*)

The authors propose a teacher-student framework based on LLMs that serves as a form of knowledge distillation. First, the large teacher model (Qwen3 72B) is prompted to produce a rationale based on the input text and the corresponding multi-label annotation. Then, a smaller (Qwen3 32B) model is fine-tuned in two phases. During the first phase, the student model learns to approximate the target rationale generated by the teacher, while at the second stage, the student model is fine-tuned to directly predict the persuasion technique labels. Furthermore, the authors employ a straightforward ensembling strategy during inference, aggregating multiple predictions for the same input sample into a voting mechanism to determine the final label. The authors used supplementary training data from the previous edition of shared tasks on persuasion techniques—CLEF-2024 and SemEval-2023.

FactUE [ST1] (Księżniak et al., 2025) (Keywords: *XLM-RoBERTa, fine-tuning, GPT-4o, embeddings, Jina, E5*)

The authors propose two approaches for building binary classifiers to recognize persuasion techniques, both leveraging multilingual transformer models. The first approach involves training data debiasing: they use GPT-4o to rewrite training samples annotated with persuasion techniques, neutralizing the persuasive style in the annotated fragments. These original and neutralized text pairs were used to fine-tune binary classifiers in a multitask setup, employing XLM-RoBERTa models. The second approach centers on "walking embeddings," where classifiers are trained on representations that capture how sentence embeddings evolve as each word is added. For this, the authors utilize two embedding models: Jina (jinaai/jina-embeddings-v3) and E5 (intfloat/multilingual-e5-large).

FactUE-ST2 [ST2] (Sawiński et al., 2025) (Keywords: *LLaMA 3.1, DeepSeek-R1, GPT-4.1-mini, Data augmentation, Prompt Engineering, Zero-shot, Fine-tuning, Automatic data generation*)

The authors propose a multi-task approach with 25 binary classification problems, one for each persuasion technique. They experiment with LlaMA3.1, DeepSeek-R1, and GPT-4.1-mini in a zero-shot setting, and GPT-4.1-mini with supervised fine-tuning using self-generated annotations by leveraging rationales on the gold labels of the training dataset combined with original text input. Additionally, the authors experimented with definition refinement, where GPT-4.1-mini was asked to produce a refined definition for each persuasion technique in a multi-step prompting process, which resulted in a significant performance gain.

GradientFlush [ST1, ST2] (Senichev et al., 2025) (Keywords: *XLM-RoBERTa, SlavicBERT,*

Data Augmentation, Fine-tuning)

The authors first enrich their training data with previous shared task editions, namely CLEF-2024 CheckThat! Task 3, leveraging data samples in English, Russian, and Polish. Furthermore, they generate synthetic data by translating English and German texts to Russian, Slovenian, and Croatian using OpenAI's GPT-4.1. They fine-tune two multilingual transformers – XLM-RoBERTa and Slavic-BERT. For Subtask 1, they only use Slavic-BERT, while for Subtask 2, they experiment with both Slavic-BERT and XLM-RoBERTa. Finally, they calibrate thresholds on the validation set separately for each language to optimize classifier performance.

INSANtive [ST1, ST2] (Wang et al., 2025) (Keywords: *Data Augmentation, GPT-4o, XLM-RoBERTa*) This paper introduces a framework for detecting persuasion techniques in five Slavic languages. The approach combines cross-lingual data augmentation, the XLM-RoBERTa architecture, and mechanisms for explanation integration – explanations are generated and then concatenated to the original text fragment. The approach achieved first rank in the Russian and Bulgarian subtasks. Key findings demonstrate that (i) larger models more effectively capture persuasive language patterns, (ii) integrating LLM-generated explanations via cross-attention mechanisms significantly improves performance, and (iii) cross-lingual augmentation effectively addresses data scarcity in low-resource languages within the same language family.

Oplot [ST1, ST2] (Loginova, 2025) (Keywords: *XLM-RoBERTa, E5, MiniLM-L12, Fine-tuning, TF-IDF*)

The authors present an approach based on fine-tuning pretrained multilingual transformer models for two tasks: binary sentence classification for subtask 1 and token-level multi-label classification for subtask 2. For subtask 1, they select the intfloat/multilingual-e5-small due to the validation set results. Interestingly, the authors perform particularly poorly on the Russian language, for which they have labeled additional news data with high-annotator agreement. For subtask 2, the authors take the token classification approach on XLM-RoBERTa, achieving rather low results. Comparing their results to proprietary large language models (LLMs) such as Claude, GPT, and Gemini, the authors demonstrate improvements in the case of few-shot models for task 1 and an overall improvement for task 2. As a baseline, the authors use TF-IDF features and SVM, achieving significantly lower results than with their system or the shared task baseline.

PSAL_NLP [ST1, ST2] (Jose and Greenstadt, 2025) (Keywords: *GPT-4o, Chain-of-thought prompting, Zero-shot, Few-shot*)

The authors present an LLM-based method, using OpenAI's GPT-4o-mini and o4-mini (the only model used for subtask 1). For subtask 1, the authors use o4-mini by prefixing each paragraph with definitions of 25 persuasion techniques, and instructing the model to output 1 of 0 based on the presence of any PRs. For subtask 2, the authors use a chain-of-thought prompt to check each paragraph against each of the 25 PTs, instructing the model to output 1 if the PT is present or 0 otherwise. The authors evaluate several prompt structures with varying amounts of contextual information and investigate performance trade-offs in terms of precision and recall.

Syntax_Squad [ST1] (Yahan et al., 2025) (Keywords: *XLM-RoBERTa, SlovakBERT, BERT-BG-WEB, Ensemble, RuBERT, SloBERTa, HerBERT*)

This paper presents an approach to detecting persuasion techniques in Slavic languages using both an extensive collection of language-specific single transformer models, like BG-BERT, RuBERT, SlovakBERT, and others, and weighted ensemble methods. It presents results only for Task 1, specifically the binary classification of the presence of persuasion in Bulgarian, Polish, Slovene, and Russian text fragments. Various pre-processing steps are applied to improve model performance. The results of the experiments show that weighted soft voting ensembles consistently outperform single models in most languages. These results demonstrate that the combination of monolingual and multilingual transformer models is effective for robust persuasion detection in low-resource Slavic languages.

UFAL4DEM [ST1, ST2] (Brückner and Pecina, 2025) (Keywords: *XLM-RoBERTa, Hierarchical classification, Hyperbolic graph convolutional networks, Data augmentation*))

The authors present an interesting take on hierarchical text classification using graph-based models embedded in hyperbolic space. Instead of treating each persuasion technique as an independent label, the authors model the label structure as a graph where each node represents a technique, and

edges reflect hierarchical relationships. Text embeddings are extracted using the domain-adapted multilingual transformer XLM-R-parla, and these are projected into the node space of the graph. The classification task is then treated as a node classification problem within this graph. The results do not outperform a standard non-hierarchical XLM-RoBERTa classifier trained on the same data, but experiments show improvements when using hyperbolic geometry compared to their Euclidean counterparts.