Multilabel Classification of Persuasion Techniques with self-improving LLM agent: SlavicNLP 2025 Shared Task

Marcin Sawiński and Krzysztof Węcel and Ewelina Księżniak
marcin.sawinski, krzysztof.wecel, ewelina.ksiezniak@ue.poznan.pl
Poznań University of Economics and Business
Poznań, Poland

Abstract

We present a system for the SlavicNLP 2025 Shared Task on multilabel classification of 25 persuasion techniques across Slavic languages. We investigate the effectiveness of in-context learning with one-shot classification, automatic prompt refinement, and supervised fine-tuning using self-generated annotations. Our findings highlight the potential of LLM-based system to generalize across languages and label sets with minimal supervision.

1 Introduction and related work

Identifying persuasion techniques in text is essential for analyzing political and social discourse (Piskorski et al., 2023). This paper presents our system for the SlavicNLP 2025 Shared Task (Piskorski et al., 2025), which addresses multilabel classification of 25 persuasion techniques. Prior work (Piskorski et al., 2023, 2024; Hasanain et al., 2024b; Sawiński et al., 2023) shows the dominance of encoder-only language models, outperforming decoder-only model despite having much fewer parameters.

Alongside advancements in foundation models, the field has recently placed increasing emphasis on enhancing test-time compute for large language models (LLMs) (Snell et al., 2024) and developing diverse model adaptation strategies.

The study primarily aimed to evaluate the performance of decoder-only LLMs on the persuasion detection task, building on previous work (Hasanain et al., 2024a). This task was broken down into 25 binary classification problems, each corresponding to a specific persuasion technique defined in subtask 2 (Piskorski et al., 2025). The system was designed to be adaptable, enabling iterative refinement and self-improvement based on analysis of failed predictions. This study addressed three research questions:

RQ1: What is the performance of one-shot multilabel classification for persuasion detection using decoder-only LLMs with basic prompts derived from the original task definition?

RQ2: Can LLMs automatically generate new task definition leading to better classification performance?

RQ3: Would fine-tuning model on automatically generated annotations improve performance of persuasion detection?

2 Dataset

The dataset provided by organizers consisted of three parts *Train*, *Trial* and *Test*. We moved random 30% of examples from *Train* and merged it with *Trial* to form new split called *Dev* for validation. Deduplication was performed before splitting dataset to prevent leakage between new *Train* and *Dev* splits.

The dataset was further preprocessed using subtask 2 annotation files. First, by splitting input files into text fragments denoted with start and end location within the file. Second, 25 binary labels column were assigned for each persuasion technique. Third, 25 annotated text excerpts columns were created for each persuasion technique appearance within text as marked in the spans files.

3 Methods

3.1 Model selection

A review of available models and providers was conducted with respect to performance, resource requirements, and cost, based on the results of a test sample run against various self-hosted and managed models.
Output: A shortlist of models selected for experimentation.

3.2 Baseline results generation

Baseline classification was performed using LLM in-context learning (Dong et al., 2024). The

Proceedings of the 10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025), pages 231–253

prompting approach combined one-shot learning (Brown et al., 2020) with Chain-of-Thought prompting (Wei et al., 2023), instructing the model to generate an explanation of the classification before providing the final verdict. The model's context included a basic definition of the technique.
Output: Preliminary performance ranking of models and prompt structures.

3.3 Identification of definition shortcomings

Persuasion technique definitions are central to the task description within the prompt. We framed definition optimization as a natural language program synthesis task, following the automatic prompt engineering approach (Zhou et al., 2023). Similar to the self-reflection paradigm (Renze and Guven, 2024), we focused on improving instructions for failed predictions. We queried LLM to generate improved definitions that would help generate correct outputs for misclassified examples in the *Train* dataset.
Output: A list of candidate definition enhancements.

3.4 Definition refinement

We consolidated enhancement candidates into concise, improved formulations of persuasion technique definitions using an iterative processing with LLM.
Output: Updated definitions.

3.5 Correct reasoning outputs generation and model fine-tuning

In-context learning often results in lower accuracy and higher inference costs compared to fine-tuning (Liu et al., 2022). We created fine-tuning dataset by querying an LLM to explain the ground-truth labels for each example in the *Train* dataset, specifying whether and how the given persuasion technique was applied. We performed supervised fine-tuning using this data and tested the impact on accuracy but did not attempt to reduce prompt length to improve inference efficiency.
Output: A supervised fine-tuning datasets and fine-tuned models.

3.6 Final evaluation

We run a second round of classification using updated prompts or fine-tuned models.
Output: Final performance ranking of models and prompt variants.

4 Experiments

4.1 Model selection

We assumed constrains for model selection:

- *cost* – the inference below 50 USD and fine-tuning below 50 USD per model,
- *resource requirements* – model for inference loaded on a single 11GB GPU,
- *inference time* – above 50 tokens per second,
- *inference quality* – above 90% of responses correctly formulated and adhering to expected JSON output.

We evaluated a range of chat-tuned models – both with and without reasoning-specific fine-tuning including LLaMA 3 and 3.3 (ranging from 1B to 70B parameters), LLaMA 4 (109B), DeepSeek-R1 variants (1.5B to 14B, based on LLaMA and Qwen), Gemma 3 (1B to 12B), Mistral 7B, and GPT-4o and 4.1 in their base, mini, and nano variants.

Larger models were excluded due to hardware limitations and their limited practical utility at current price points. We also evaluated multiple quantization strategies and inference backends, including vLLM, Ollama, Llama.cpp, and Hugging Face Transformers.

Three models were ultimately selected for the experiments: **LLaMA 3.1 8B**, **DeepSeek-R1 8B** (based on LLaMA 3.1 8B but augmented with test-time reasoning), and **GPT-4.1-mini**. While OpenAI has not disclosed the exact specifications of GPT-4.1-mini, it is speculated to have between 7B and 9B parameters, making it comparable in scale to the other two. For self-hosted models, we used 4-bit grouped quantization and the Ollama backend achieving throughput of 300 tokens per second with four NVIDIA GeForce RTX 2080 Ti GPUs. The approximate cost for fine-tuning GPT-4.1-mini was 35 USD and 40 USD for inference.

4.2 Baseline results generation

The baseline results were obtained by querying the LLM with one-shot prompts for binary classification of each persuasion technique.

A single prompt template was employed across all techniques, with placeholders for the technique name, definition, example, and actual input text substituted for each query.

The original prompts (later denoted as **v1 prompts**) were based on the descriptions and examples of persuasion techniques provided by the organizers[1].

The LLMs were queried in chat mode using message chains composed of a system message followed by a user message.

The system message included:

- a task description and formatting instructions,
- the definition of the persuasion technique,
- an input-output example.

The user message contained only the raw input text to be classified.

The assistant message (i.e., the expected output) was a JSON object with two fields:

- *explanation*: reasoning as to whether the persuasion technique is present in the input and why,
- *verdict*: a boolean value indicating the presence or absence of the technique.

The *explanation* field was introduced not only as a tool for Chain-of-Thought prompting, but also to capture the model's reasoning prior to issuing a verdict, thereby enabling analysis of potential errors and guiding the refinement of the definition.

LLM outputs were validated on the fly. If a response failed to comply with the expected JSON schema, the prompt was resent using a different sampling seed or temperature setting to encourage schema adherence.

Several alternative prompt templates and output formats were tested, including multi-label classification within a single prompt, additional output fields, and confidence scoring. However, these variants frequently resulted in malformed outputs and were therefore discarded.

Appendix B.1 provides the actual prompt template used, along with the *v1* technique definitions presented in Appendix C.

4.3 Supporting Dataset Generation

Two supporting datasets were generated by processing the *Train* split with GPT-4.1-mini, once for each of the 25 persuasion techniques:

- **Correct reasoning outputs**: The model was instructed to generate a rationale justifying the gold label. These outputs were combined with the *Train* dataset to form complete message chains for supervised fine-tuning. System and user messages identical to those in the classification task, were followed by an assistant message containing a JSON object with the generated explanation and gold label.
- **Enhancement candidates**: In cases where the model's prediction differed from the gold label, it was prompted to identify shortcomings in the persuasion technique's definition and propose a revised version to improve prediction accuracy, especially in edge cases.

The prompt included:

- a task and output format description,
- the persuasion technique definition,
- the input text and its gold label,
- text spans illustrating the use of the technique.

The actual prompt template used is listed in Appendix B.2.

4.4 Definition Refinement

Enhancement candidates for each persuasion technique were consolidated using GPT-4.1-mini to produce a revised set of 25 definitions (later denoted as **v2 prompts**). The consolidation prompt included:

- the base persuasion technique definition,
- suggested updates,
- guidelines for integrating the suggestions into a refined definition.

The process was iterative: the first iteration used the original definition as the base, while subsequent iterations used the refined output from the previous step as the new base.

The intention of the Definition Refinement was not to alter the definitions per se, but to guide the model on edge cases and improve alignment between the definitions and the annotations. Only in cases of erroneous annotations or LLM errors could the definition be unintentionally skewed in the wrong direction.

The prompt template used is provided in Appendix B.3, and the resulting updated definitions are presented in Appendix C denoted as *v2*.

[1] http://bsnlp.cs.helsinki.fi/PT-TAXONOMY.pdf

4.5 Supervised Fine-Tuning

Three fine-tuning datasets were generated with varying ratios of negative (majority) to positive class: 1:1, 2:1, and 3:1. Baseline results revealed significantly lower precision than recall. Since earlier attempts to derive a reliable confidence or probability score were unsuccessful, it was not possible to control the precision–recall trade-off post hoc.

To address this, imbalanced training sets were intentionally constructed to reduce the model's tendency toward false positives.

Fine-tuning was performed on GPT-4.1-mini using the OpenAI service for 3 epochs, with a batch size of 4 and an LR multiplier of 2, using a total of 5,476,698 tokens.

4.6 Additional Classification Runs

In the final phase of the experiment, additional classification runs were executed on the *Dev* dataset using updated persuasion definitions, fine-tuned models, or both, and were compared with baseline results.

5 Results

5.1 Evaluation procedure

Results were processed using the official evaluation script, extended with two additional features. First, we added the more forgiving Hamming loss metric alongside exact match accuracy to better capture partial correctness. Second, we introduced functionality to compute metrics on the combined dataset across all languages, avoiding the averaging of precomputed scores. This approach aligned with our decision to use the same model and its adaptations across all target languages.

5.2 Baseline results

Results obtained with the original *v1* prompts and not fine-tuned models across all languages showed that GPT-4.1-mini consistently outperformed both LLaMA 3.1 8B and DeepSeek-R1 8B on all evaluation metrics, including Hamming loss, precision, recall, and F1 score (both micro- and macro-averaged). No clear second-best model emerged: while DeepSeek achieved a lower Hamming loss, it lagged behind LLaMA in precision, recall, and F1 score aggregated across all classes (see Figure 1).

All models exhibited a tendency toward false positives, reflecting a consistent bias toward higher recall over precision, which severely impacted the F1 score (0.78 vs. 0.15 for GPT, 0.64 vs. 0.11 for DeepSeek, and 0.76 vs. 0.11 for LLaMA).

5.3 Fine-tuning results

No direct qualitative or quantitative analysis was conducted on the automatically generated explanations used in the fine-tuning dataset. Their quality was assessed solely through the performance of the downstream task—that is, the results achieved by the fine-tuned model.

Fine-tuning of the best model—GPT-4.1-mini—brought mixed results. While it improved the micro F1 score compared to the not fine-tuned version, it resulted in a lower macro F1 score. A clear drop in recall was observed, accompanied by a substantial improvement in precision. Accuracy and Hamming loss were also exceptionally good.

5.4 Definition refinement results

The identification of *v1* technique definition shortcomings involved sending 15,491 queries to GPT-4.1-mini and resulted in the generation of 1,038 technique definition enhancement candidates, unevenly distributed across all techniques—50% were generated for just three techniques: Repetition, Slogans, and Name Calling-Labeling.

During the persuasion technique definition refinement stage, the enhancement candidates were aggregated into new *v2* definitions that were significantly longer, increasing the average length from 564 to 1,733 characters. Qualitative analysis showed that the additional text tended to expand rather than constrain the definitions, and many repetitions were observed.

In the second round of classification, we collected predictions using the updated *v2* prompts, along with outputs from an additional GPT-4.1-mini model fine-tuned on the gold labels and reasoning dataset with a 2:1 ratio of negative examples.

The use of updated *v2* prompts improved both micro and macro F1 scores across all models—except LLaMA, which experienced a decline in performance. Depending on the model, the improvement was driven by gains in precision, recall, or both.

Analysis of the results by language revealed that the task was easier for some languages than for others, regardless of the prompt version or model used, with results for Slovenian and Polish outperforming those for Russian and Bulgarian (see Figures 2, 3, and 4 in the Appendix).

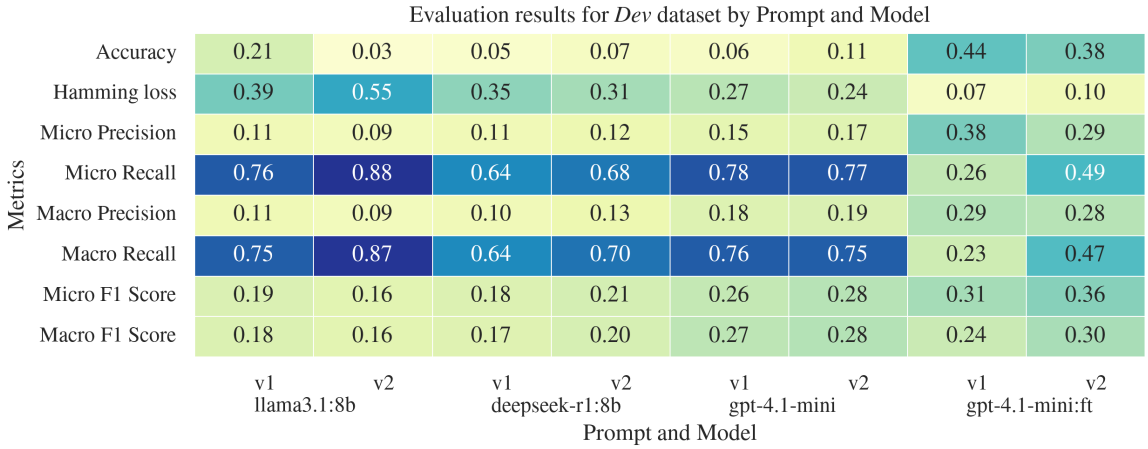

Figure 1: Evaluation results for all languages on the *Dev* dataset.

Analysis of the results by persuasion technique revealed that, for most classes, *v2* prompts reduced precision but improved recall and F1 score compared to *v1* prompts (see Figures 6, 7, and 8 in the Appendix).

The F1 scores computed for each technique vary significantly, regardless of the model used. This variation may be attributed not only to the inherent difficulty of classifying specific techniques but also to class imbalance and differences in distribution across languages. An in-depth analysis of the results for all 25 persuasion techniques is beyond the scope of this paper.

The fine-tuned GPT model outperformed all non–fine-tuned models on most metrics, even without using the *v2* prompts. Although recall dropped significantly (from 0.78 to 0.26), a substantial increase in precision (from 0.15 to 0.38) ensured a strong F1 score. When combined with the *v2* prompts, the fine-tuned GPT model achieved the best overall results. The more detailed instructions in the updated prompts helped the model partially recover its recall, with only a modest reduction in precision.

5.5 Task leaderboard analysis

The best-performing model (GPT-4.1-mini fine-tuned with improved persuasion technique definitions) tested on an unseen dataset compares favorably with other submissions to the SlavicNLP 2025 Shared Task.

In terms of macro F1 score, the model achieved the highest score for Croatian, the second-best for Bulgarian, Polish, and Slovenian, and the third-best for Russian. For micro F1 score, it ranked second for both Croatian and Slovenian, placing the model among the top three performers in 7 out of 10 evaluation categories.

Its advantage in macro F1 score—particularly in less-represented languages—may indicate a superior ability to generalize to out-of-distribution data or to perform well when training resources are limited.

Notably, some evaluation data were submitted after the official deadline and are therefore not reflected in the official competition ranking.

6 Conclusions

Our experiments demonstrate the superiority of the fine-tuned GPT-4.1-mini model. Refined definitions significantly improved model precision without severely compromising recall, resulting in high F1 scores across multiple Slavic languages. Results underscore the effectiveness of supervised fine-tuning with generated explanations and iterative self-improvement strategies in LLM systems, specifically automated prompt refinement. This approach offers promising directions for future research, especially in multilingual settings or scenarios with limited training resources.

References

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. *Preprint*, arXiv:2005.14165.

Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Jingyuan Ma, Rui Li, Heming Xia, Jingjing Xu, Zhiyong Wu, Tianyu Liu, Baobao Chang, Xu Sun, Lei Li, and Zhifang Sui. 2024. A survey on in-context learning. *Preprint*, arXiv:2301.00234.

Maram Hasanain, Fatema Ahmad, and Firoj Alam. 2024a. Can GPT-4 identify propaganda? annotation and detection of propaganda spans in news articles. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2724–2744, Torino, Italia. ELRA and ICCL.

Maram Hasanain, Md. Arid Hasan, Fatema Ahmed, Reem Suwaileh, Md. Rafiul Biswas, Wajdi Zaghouani, and Firoj Alam. 2024b. Araieval shared task: Propagandistic techniques detection in unimodal and multimodal arabic content. *Preprint*, arXiv:2407.04247.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. 2022. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Preprint*, arXiv:2205.05638.

Jakub Piskorski, Dimitar Dimitrov, Filip Dobranić, Marina Ernst, Jacek Haneczok, Ivan Koychev, Nikola Ljubešić, Michał Marcińczuk, Arkadiusz Modzelewski, Ivo Moravski, and Roman Yangarber. 2025. SlavicNLP 2025 Shared Task: Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Jakub Piskorski, Alípio Jorge, Maria da Purificação Silvano, Nuno Guimarães, Ana Filipa Pacheco, and Nana Yu. 2024. Overview of the clef-2024 checkthat! lab task 3 on persuasion techniques.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multilingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

Matthew Renze and Erhan Guven. 2024. The benefits of a concise chain of thought on problem-solving in large language models. In *2024 2nd International Conference on Foundation and Large Language Models (FLLM)*, page 476–483. IEEE.

Marcin Sawiński, Krzysztof Węcel, Ewelina Paulina Księżniak, Milena Stróżyna, Włodzimierz Lewoniewski, Piotr Stolarski, and Witold Abramowicz. 2023. Openfact at checkthat! 2023: head-to-head gpt vs. bert-a comparative study of transformers language models for the detection of check-worthy claims. In *CEUR Workshop Proceedings*, volume 3497.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. 2024. Scaling llm test-time compute optimally can be more effective than scaling model parameters. *Preprint*, arXiv:2408.03314.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models. *Preprint*, arXiv:2201.11903.

Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. 2023. Large language models are human-level prompt engineers. *Preprint*, arXiv:2211.01910.

A Detailed evaluation results

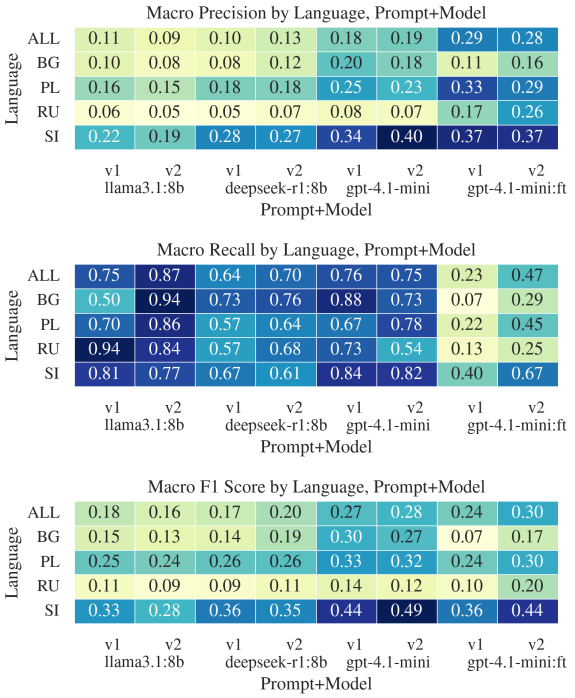

Figure 2: Macro-averaged metrics by language for *Dev* dataset.

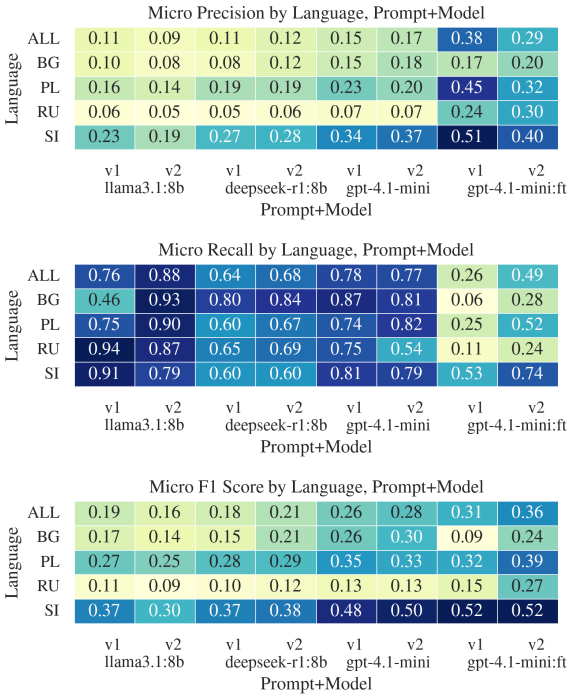

Figure 4: Micro-averaged metrics by language for *Dev* dataset.

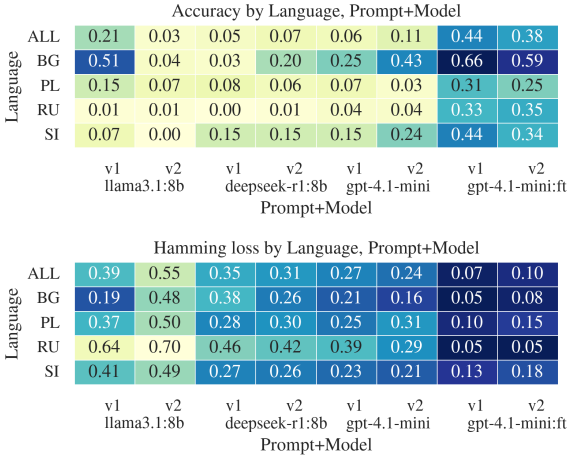

Figure 3: Accuracy and Hamming loss metrics by language for *Dev* dataset.

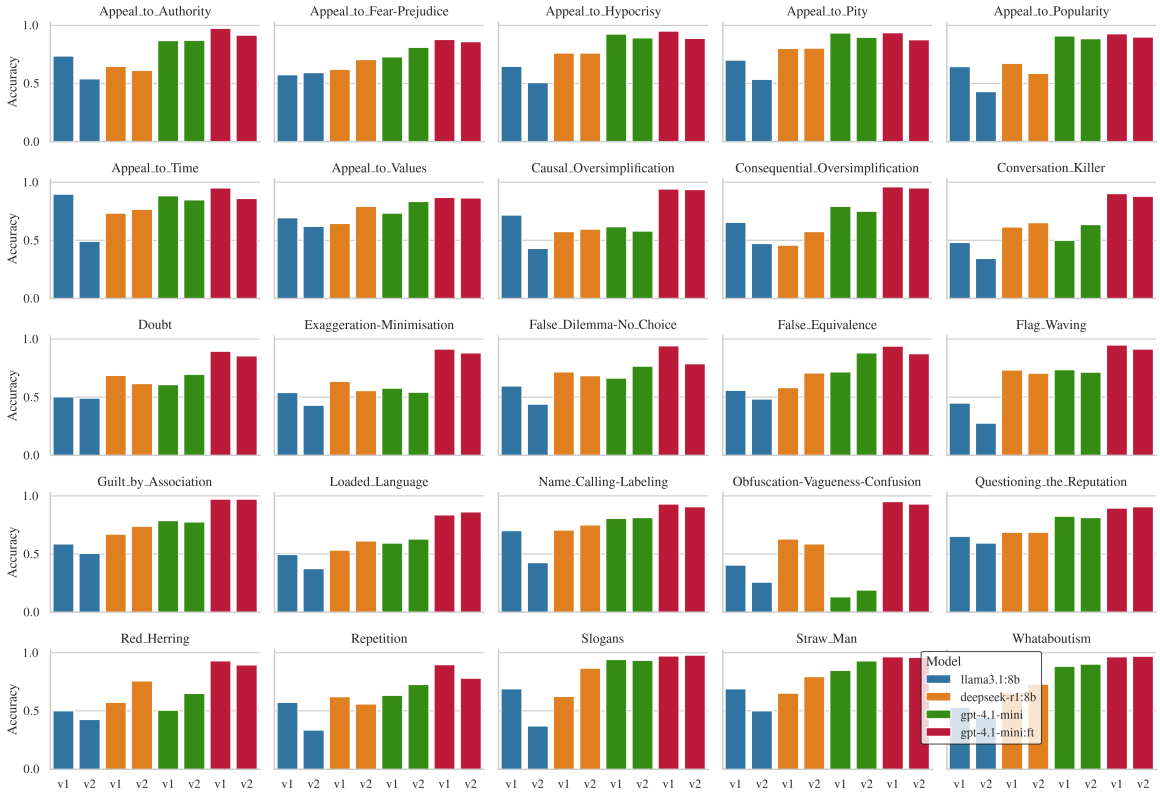

Figure 5: Accuracy per class for all languages for *Dev* dataset.

Figure 6: Precision per class for all languages for *Dev* dataset.

Figure 7: Recall per class for all languages for *Dev* dataset.

Figure 8: F1 Score per class for all languages for *Dev* dataset.

B Prompt templates

B.1 Persuasion technique detection prompt

system:You are a classifier for persuasion technique called <TECHNIQUE_NAME>.

Input: a chunk of text

Task: Detect if a specific technique called <TECHNIQUE_NAME> was explicitly used in the input.

Output: JSON file with 2 fields:
- description - describes if and how <TECHNIQUE_NAME> technique was used.
- verdict - output value 'true' if you have high confidence that <TECHNIQUE_NAME> technique was used in the input text and 'false' otherwise.

Follow the definition of the <TECHNIQUE_NAME> technique:
<TECHNIQUE_DEFINITION>

Output format:
{"description": "if and how <TECHNIQUE_NAME> technique was used.",
"verdict": true|false}

Do not add any other information. Output **only** valid JSON.
user:<TEXT>

B.2 Supporting dataset generation prompt

system: You are a helpful assistant that explains a verdict of an expert regarding use of a specific persuasion called technique <TECHNIQUE_NAME> in an input text.

Inputs:
- an input text
- a binary verdict of an expert (the technique was used or not)
- excerpts from the input text indicated by the expert that show the application of <TECHNIQUE_NAME> technique.

Tasks:
- Write a very concise explanation if and how <TECHNIQUE_NAME> was explicitly used in the input text comparing the technique definition and excerpts.
- If the expert verdict cannot be explicitly derived from definition **NEVER** change the verdict. Instead improve the definition in such a way that the verdict can be explicitly derived from the updated definition.

Output: JSON file with 2 fields:
- explanation - mandatory - describes if and how <TECHNIQUE_NAME> technique was used.
- updated_definition - optional - improved definition that guides in edge cases like the one provided in input.

<TECHNIQUE_NAME> technique definition: <TECHNIQUE_DEFINITION>
Do not add any other information. Output **only** valid JSON.
user: Input text: <TEXT>
Expert verdict: <GOLD_LABEL>
Excerpts supporting verdict: <TEXT_FROM_SPANS>

B.3 Persuasion technique definition update prompt

system: You are a helpful assistant that improves a definition for <TECHNIQUE_NAME> persuasion technique used for automated labeling. You will receive:
- A **base definition** of <TECHNIQUE_NAME> persuasion technique.
- An **update suggestion**, which may introduce new elements or clarify edge cases.

Your task is to revise the base definition by integrating meaningful, non-duplicate additions from the suggestion. Guidelines:
- Identify elements in the update that are **not already present** in the base.
- Add only **new and relevant elements** to the base definition.
- **Do not remove** any existing important elements from the base, even if absent in the update.
- If the update merely rephrases elements already covered in the base, **do not include** them again.
- Avoid **duplication**.
- It is acceptable to **extend** the definition with clarifications, examples, restrictions, or edge-case descriptions if they improve accuracy.
- Ensure the final definition is **reasonably concise**, but complete.
- Aim for the definition to be **mutually exclusive

and collectively exhaustive**.

Output:
- Return only the **updated definition** as plain text.
- Do **not** include any additional commentary, explanation, or formatting.
user: Base definition: `<definition_base>`
Definition update suggestion: `<definition_update>`

C Persuasion techniques definitions

C.1 Name Calling - Labeling

v1: a form of argument in which loaded labels are directed at an individual or a group, typically in an insulting or demeaning way. Labelling an object as either 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 in a way also a manipulative wording, as it is used at the level of the nominal group rather than being a full-fledged argument with a premise and a conclusion. For example, in the political discourse, typically one is using 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 the Loaded Language technique, is that it is only concerned with the characterization of the subject.
v2: Name Calling - Labeling is a form of argument in which loaded labels are directed at an individual or a group, typically in an insulting or demeaning way, to evoke a strong emotional response, without providing factual support. This technique involves characterizing a subject with qualitative judgments that evoke fear, hatred, or desirability, often disregarding factual evidence. It is characterized by manipulative wording that targets the essence of the subject being characterized, often using adjectives and nouns as labels that refer to political orientation, personal characteristics, opinions, and associations, in a derogatory manner. In political discourse, it often employs labels referring to political orientation, opinions, personal characteristics, and associations, as well as insults. Name Calling - Labeling reinforces social divisions and biases by framing the labeled individuals or groups in a negative light, creating an in-group versus out-group dynamic. It exploits existing prejudices or stereotypes to enhance its impact and influence the audience's attitudes and beliefs without engaging in rational discourse. Additionally, it can further entrench the audience's views and potentially lead to polarization, as it simplifies complex issues into binary categories that discourage nuanced understanding. The absence of such labeling in a text indicates that this technique has not been employed. This technique specifically focuses on the impact of the labels rather than the content of the argument itself, often in a manipulative manner, and can be particularly effective in shaping public perception and opinion in a way that aligns with the speaker's agenda. It is distinct from Loaded Language as it emphasizes characterization rather than full arguments and involves qualitative judgments that disregard facts, distinguishing it from other forms of loaded language.

C.2 Guilt by Association

v1: Attacking the opponent or an activity by associating it with another group, activity, or concept that has sharp negative connotations for the target audience. The most common example, which has given its name in the literature (i.e. Reduction ad Hitlerum) to that technique is making comparisons to 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 to any individual, group, or event in the presence or in the past, which has or had an unquestionable negative perception (e.g., was considered a failure), or is depicted in such way.
v2: Attacking the opponent or an activity by associating it with another group, activity, or concept that has sharp negative connotations for the target audience, often by implying that the opponent's actions lead to negative consequences or suggesting that their actions result in harmful outcomes. The most common example, which has given its name in the literature (i.e. Reduction ad Hitlerum) to that technique, is making comparisons to 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 to any individual, group, or event in the presence or in the past, which has or had an unquestionable negative perception (e.g., was considered a failure), or is depicted in such a way.

C.3 Casting Doubt

v1: Casting doubt on the character or the personal attributes of someone or something in order to question their general credibility or quality, instead of using a proper argument related 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 (probably) resulting in not achieving the planned goals.
v2: Casting doubt on the character or the personal attributes of someone or something in order to question their general credibility or quality, instead of using a proper argument related to the topic. This can be done by highlighting inconsistencies in their actions or statements, referencing failures or inefficiencies in processes or systems, and discussing the target's professional background or actions that suggest incompetence or negligence, thereby discrediting their argument or proposals. Casting doubt can also be achieved by referring to unsuccessful actions or events carried out or planned by the entity, or by making strong assertions that challenge the effectiveness or intentions of a policy or action. Additionally, it can involve expressing skepticism about the effectiveness or quality of a process or outcome. The technique is characterized by explicit statements or implications that challenge the integrity or reliability of the subject, often leading the audience to question the validity of the subject's claims or arguments. This includes explicit references to the target's credibility or past failures, as well as actions or events that are perceived as likely to result in failure to achieve planned goals. Furthermore, it often involves contrasting positive attributes of one group with negative attributes of another to create skepticism about the latter. It often includes asserting that the individual lacks knowledge or awareness of critical issues, which further undermines their credibility. The technique is characterized by the presence of specific claims or insinuations that undermine credibility, often focusing on perceived shortcomings or failures of the target, and is marked by negative implications or insinuations about the target's reliability or effectiveness. This can also include highlighting inconsistencies in their actions or priorities, such as their focus on less significant issues while ignoring more pressing problems, and suggesting that others have overlooked or ignored critical information, thereby questioning their integrity or intentions. Additionally, it can involve questioning the specifics or details of a claim, such as by asking for missing information, which implies that the original statement may not be trustworthy. It can also include questioning the engagement or presence of an audience in discussions, implying a lack of credibility in their participation, and expressing uncertainty about how to engage with the audience, thereby questioning their interest and credibility. The technique can also involve highlighting the minority status of dissenting opinions to undermine their legitimacy. The technique is characterized by an explicit challenge to the credibility of the source or the information presented, and it can be employed as a tactic to divert attention from the actual argument by focusing on the perceived flaws of the individual or entity involved.

C.4 Appeal to Hypocrisy

v1: The target of the technique is attacked on its reputation by charging them with hypocrisy or inconsistency. This can be done explicitly by calling out hypocrisy directly, or more implicitly by underlying the contradictions between different positions that were held or actions that were done in the past. A special way of calling out hypocrisy is by telling that someone who criticizes you for something you did, also did it in the past.
v2: The target of the technique is attacked on its reputation by charging them with hypocrisy or inconsistency, either explicitly by calling out hypocrisy directly, or implicitly by highlighting contradictions between their past statements or actions and their current claims. This includes showing how similar actions are treated differently based on the target's alignment with certain interests. A special way of calling out hypocrisy is by stating that someone who criticizes you for something you did also did it in the past, thereby showing inconsistency in their stance, or by contrasting current claims with past actions.

This includes instances where someone criticizes another for a behavior they themselves have exhibited, thereby undermining their credibility and moral authority in the discussion. It particularly applies when the critic has previously engaged in the same behavior they are condemning or when their actions contradict their stated goals. Additionally, it can involve highlighting specific instances where the individual claims to uphold certain principles or standards but acts in a manner that directly contradicts those claims. The technique is also evident when a person's claims are directly contradicted by factual evidence or statistics, further emphasizing the inconsistency in their position. Furthermore, the technique often involves questioning the memory or awareness of the audience regarding the target's past actions, thereby manipulating perceptions of the target's credibility. This technique is particularly effective when the target's criticisms are juxtaposed with their own similar past behaviors, reinforcing the perception of hypocrisy. It also encompasses situations where the critic's own past behaviors are revealed, thereby illustrating their inconsistency and further attacking their credibility.

C.5 Questioning the Reputation

v1: This technique is used to attack the reputation of the target by making strong negative claims about it, focusing specially on undermining its character and moral stature rather than relying on an argument about the topic. Whether the claims are true or false is irrelevant for the effective use of this technique. Smears can be used at any point in a discussion. One particular way of using this technique is to preemptively call into question the reputation/credibility of an opponent, before he had any chance to express himself, therefore biasing the audience perception. Hence, one of the name of that technique is "poisoning the well." The main difference between Casting Doubt and Questioning the reputation technique is that the former focuses on questioning the capacity, the capabilities, and the credibility of the target, while the latter targets undermining the overall reputation, moral qualities, behaviour, etc.

v2: This technique is used to attack the reputation of the target by making strong negative claims about it, focusing especially on undermining its character and moral stature rather than relying on an argument about the topic. Whether the claims are true or false is irrelevant for the effective use of this technique. Smears can be used at any point in a discussion. One particular way of using this technique is to preemptively call into question the reputation/credibility of an opponent, before they have any chance to express themselves, therefore biasing the audience's perception. The main difference between Casting Doubt and Questioning the Reputation technique is that the former focuses on questioning the capacity, capabilities, and credibility of the target, while the latter targets undermining the overall reputation, moral qualities, behavior, etc. This technique is characterized by explicit negative claims or insinuations about the target's character or integrity, which can manifest as personal attacks, character assassination, or the spreading of rumors. It often aims to create a lasting negative impression that can influence the audience's perception beyond the immediate context of the discussion and may involve the use of emotionally charged language to elicit a strong reaction from the audience. Additionally, this technique often involves direct questioning of the actions or integrity of the target to provoke doubt in the audience's mind, and it can also include the use of anecdotal evidence or selective information to reinforce negative perceptions. Furthermore, it often involves linking the target to negative behaviors or outcomes to damage their standing, and it may exploit existing biases or stereotypes to enhance the effectiveness of the attack. The technique is characterized by direct attacks on the target's character or moral standing, emphasizing the intent to damage the target's reputation rather than engage in substantive debate. This technique often involves rhetorical questions that challenge the target's integrity or actions.

C.6 Flag Waving

v1: Justifying or promoting an idea by exhaling 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 group it applies to might 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 fully undue and is usually based on the presumption that the recipients already have 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 the form of well constructed argument, by simply making mentions that resonate with the feeling of a particular group and as such setting up a context for further arguments.

v2: Flag Waving persuasion technique involves justifying or promoting an idea by appealing to the pride or benefits of a specific group, often through emotional manipulation rather than logical argumentation. This technique is characterized by references that resonate with the feelings of a particular group, such as national pride, race, gender, or political preference, and aims to evoke a sense of belonging or identity related to that group. It can manifest in various forms, including direct statements of pride, expressions of gratitude, or actions that are framed to evoke a sense of loyalty or emotional connection to a group, as well as contextual references that highlight their struggles or needs. The connection to nationalism, patriotism, or benefit for an idea, group, or country might be fully undue and is usually based on the presumption that the recipients already have certain beliefs, biases, and prejudices about the given issue. It can be seen as an appeal to emotions instead of logic of the audience, aiming to manipulate them to win an argument. Additionally, it emphasizes the audience's identity and values, further enhancing the emotional connection to the message being conveyed. This technique often relies on emotional appeals that resonate with the audience's existing beliefs, biases, or prejudices and can include both direct statements of pride and indirect references that evoke group identity. It typically involves references to group identity, such as nationality, race, or political affiliation, to manipulate the audience's emotions rather than presenting logical arguments, reinforcing their emotional responses to the message. Furthermore, the Flag Waving technique sets a context for further arguments, leveraging the emotional resonance to strengthen the overall persuasive impact. It also involves invoking feelings of nationalism and patriotism to manipulate the audience's perception and arguments. The stereotypical example would be national pride; however, the target group it applies to might be any group, including those related to race, gender, or political preference, and it can appear in both structured arguments and casual mentions that evoke group feelings, particularly by highlighting the impact on the group's well-being. This technique is characterized by emotional resonance rather than logical argumentation, aiming to manipulate the audience's feelings through references to group identity, nationalism, or shared values. It is typically aimed at manipulating the audience's beliefs or biases through emotional appeals that resonate with the feelings of a particular group, emphasizing the audience's sense of identity and collective sentiment. Additionally, it can include references to the benefits for a community, thereby resonating with the audience's feelings of national or group identity, and highlighting the pride of the group or the advantages that the idea may bring to them.

C.7 Appeal to Authority

v1: a weight is given to an argument, an idea or information by simply stating that a particular entity considered as an authority is the source of the information. The entity mentioned as an authority may, but does not need to be, an actual valid authority in the domain-specific field to discuss a particular topic or to be considered and serve as an expert. What is important, and makes it different from simply sourcing information, is that the tone of the text indicates that it capitalizes on the weight of an alleged authority in order to justify some information, 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.

v2: The Appeal to Authority technique involves giving weight to an argument, an idea, or information by stating that a recognized authority supports the information or claim. This technique is characterized by the text's tone indicating reliance on the authority's credibility to justify claims. It is important that the authority is explicitly mentioned and that the argument relies on their status to lend support to the information presented, rather than making general statements about an individual's qualities without citing their authority. The entity mentioned as an authority may, but does not need to be, an actual valid authority in the domain-specific field to discuss a particular topic or to be considered and serve as an expert. What distinguishes this technique from simply sourcing

information is that the tone of the text capitalizes on the weight of the alleged authority to justify some information, claim, or conclusion. This includes referencing the opinions or experiences of individuals in relevant fields, such as professionals or experts, to support claims. The technique is characterized by the use of authoritative statements to bolster arguments, often without critical examination of the authority's validity. The tone of the text should indicate that the authority's status is leveraged to support the argument, rather than simply presenting information or statistics without context. The technique is explicitly identified when the authority is presented as a source that supports the argument being made. 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, where the author uses their own credentials or experiences to support their argument.

C.8 Appeal to Popularity

v1: This technique gives weight to an argument or idea by justifying it on the basis that allegedly "everybody" (or the vast majority) agrees with it or "nobody" disagrees with it. As such, the target audience is encouraged to gregariously adopt the same idea by considering "everyone else" as an authority, and to join in and take the course of the same action. Here, "everyone else" 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.

v2: This technique gives weight to an argument or idea by justifying it on the basis that allegedly "everybody" (or the vast majority) agrees with it or "nobody" disagrees with it. The target audience is encouraged to adopt the same idea by considering "everyone else" as an authority, and to join in and take the course of the same action. Here, "everyone else" might refer to the general public, key entities and actors in a certain domain, or even entire countries. This technique can also involve attempts to persuade the audience not to do something because "nobody else is taking the same action," thereby leveraging the fear of social exclusion or being out of step with the majority. Additionally, references to public opinion or widespread acknowledgment of an issue can also indicate the use of this technique.

C.9 Appeal to Values

v1: This technique gives weigh to an idea by linking it to values seen by the targe audience as positive. These values are presented as an authoritative reference in order to support or t 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 certain 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.

v2: This technique gives weight to an idea by linking it to values seen by the target audience as positive, such as tradition, religion, ethics, age, fairness, liberty, democracy, peace, transparency, safety, integrity, and accountability, to support or reject an argument. The values must be explicitly referenced as authoritative references to support or reject an argument, rather than merely being implied or mentioned in passing. Additionally, the appeal to values can be particularly effective when the values resonate deeply with the audience's identity or beliefs, enhancing the emotional connection to the argument. It is important that the values are not only referenced but are also relevant and significant to the audience's context, ensuring that the appeal is meaningful and impactful. The technique is distinct from Loaded Language, which involves using certain adjectives or nouns to characterize something or someone without a proper argument. Furthermore, the values should not be presented in a critical context, as this may undermine their perceived authority and relevance. The appeal to values should also avoid vague or generic references, focusing instead on specific values that hold particular significance for the audience. Additionally, the values must be presented in a positive context to reinforce their authority and relevance in supporting the argument. If values are mentioned without a clear connection to an argument, they may fall under Loaded Language, highlighting the necessity of a

direct link between the values and the argument being made. The presentation of these values should actively support or reject an argument, particularly in contexts where the audience's well-being or moral standards are at stake, rather than merely stating facts, ensuring a clear and persuasive connection. Moreover, the values should not be used solely for emotional appeals but must serve a clear purpose in the argumentation process. It should be noted that mere negative characterizations or references to values without a constructive appeal do not qualify as the Appeal to Values technique. Additionally, these values must be referenced in a way that supports or rejects an argument, rather than merely describing actions or positions, and should not be discussed in a procedural context.

C.10 Appeal to Fear - Prejudice

v1: This technique aims at promoting or rejecting an idea through the repulsion or fear of the audience towards this idea (e.g., via exploiting some preconceived judgements) or towards 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, and if there are only two alternatives that are stated explicitly, then it is used simultaneously with the False Dilemma technique.

v2: This technique aims at promoting or rejecting an idea through the audience's fear or repulsion towards that idea or its alternatives, often by exploiting preconceived judgments and suggesting dire consequences. The alternative could be the status quo, in which case the current situation is described in a frightening manner, using loaded language to create a sense of fear. It can involve describing a scary scenario related to the idea or its consequences, often using emotionally charged language that implies dire outcomes to provoke an immediate emotional response rather than a rational evaluation of the situation. Additionally, it may exploit societal prejudices or stereotypes to amplify fear, further manipulating the audience's emotional state and decision-making process. The technique often involves creating a narrative that emphasizes the dangers or negative outcomes associated with a particular group or political decision, thereby inciting fear and prejudice. It may also include questioning the audience's trust or safety regarding a particular subject. If the fear is linked to the consequences of a decision, it may overlap with Appeal to Consequences, and if only two alternatives are presented explicitly, it may also involve the False Dilemma technique. The presence of fear-inducing language or implications is essential for identifying this technique. It can be identified through language that evokes fear or repulsion regarding a concept or its consequences, and it often creates a sense of urgency, compelling the audience to act quickly based on fear rather than careful consideration. The absence of fear or negative implications in the message indicates that this technique is not being used. This technique often employs loaded language to evoke strong emotional responses, further enhancing its persuasive impact, and it may create a sense of urgency to prompt immediate action.

C.11 Strawman

v1: This technique consists in making an impression of refuting the argument of the opponent's proposition, whereas 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 misrepresentation of the argument. First, a new argument is created via the covert replacement of the original argument with something that appears somewhat related, but is actually a different, a distorted, an exaggerated, or a misrepresented version of the original proposition, which is referred to as "standing up a straw man." Subsequently, the newly created 'false argument (the strawman) is refuted, which is referred to as "knocking down a straw man." Often, the strawman argument is created in such a way that it is easier to refute, and thus, creating an illusion of having defeated an opponent's real proposition. Fighting a strawman is easier than fighting against a real person, which explains the origin of the name of this technique. In practice, it appears often as an abusive reformulation or explanation of what the opponent actually' means or wants.

v2: This technique consists in making an impression of refuting the argument of the opponent's proposition by replacing it with a false one, which is a distorted, exaggerated, or misrepresented version of the original argument. Often, this technique is referred to as misrepresentation of

the argument. First, a new argument is created via the covert replacement of the original argument with something that appears somewhat related, but is actually different. The newly created argument (the strawman) is then refuted, creating an illusion of having defeated the opponent's real proposition. This technique often appears as an abusive reformulation or explanation of what the opponent actually means or wants, particularly by reducing their argument to an oversimplified or extreme version that can be easily attacked, thereby distorting the original argument further. Fighting a strawman is easier than fighting against a real person, which explains the origin of the name of this technique. This technique requires the presence of an original argument to misrepresent; without such an argument, the technique cannot be applied.

C.12 Red Herring

v1: 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 (like a herring) can be used to divert dogs from the scent of someone they are following. A strawman (defined earlier) is also a specific type of a red herring in the way that it distracts from the main issue by painting the opponent's argument in an inaccurate light.

v2: This technique consists in diverting the attention of the audience from the main topic being discussed by introducing another unrelated topic or issue, making it difficult to address the original argument. The aim is to redirect the argument to something the person doing the redirecting can better respond to or to leave the original topic unaddressed. A clear main topic must be present for the technique to be applicable. The name of this technique comes from the idea that a fish with a strong smell (like a herring) can be used to divert dogs from the scent of someone they are following. This technique may involve introducing a topic that seems related but ultimately distracts from the original issue, thereby redirecting the focus away from the main point. A clear example of this technique would involve specific statements that shift the focus away from the main issue. A strawman (defined earlier) is also a specific type of a red herring in the way that it distracts from the main issue by painting the opponent's argument in an inaccurate light. Additionally, the red herring technique can manifest through the introduction of irrelevant information that may appear to have some connection but ultimately serves to mislead or confuse the audience regarding the original topic. It can also involve introducing a related context that does not directly address the original issue, further complicating the discussion.

C.13 Whataboutism

v1: A technique that attempts to discredit an opponent's position by charging them with hypocrisy without directly disproving their argument. Instead of answering a critical question or argument, an attempt is made to retort with a critical counter-question that expresses a counteraccusation, e.g., mentioning double standards, etc. The intent is to distract from the content of a topic and to switch the topic actually. There is a fine distinction between this technique and Appeal to Hypocrisy, introduced earlier, where 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 very related topic.

v2: A technique that attempts to discredit an opponent's position by charging them with hypocrisy without directly disproving their argument. Instead of answering a critical question or argument, an attempt is made to retort with a critical counter-question or suggestion that expresses a counteraccusation, e.g., mentioning double standards or what the opponent should have done instead. The intent is to distract from the content of a topic and to switch the topic, often by referencing past actions or failures of the opponent. There is a fine distinction between this technique and Appeal to Hypocrisy, where 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 very related topic. Whataboutism often involves emotionally charged or unrelated issues to undermine the opponent's stance and can manifest in various forms, such as shifting the focus to the opponent's past behavior or unrelated controversies. Additionally, it may

involve a false equivalence, suggesting that the opponent's actions are comparable to the issue at hand, further complicating the discourse and obscuring the original argument. This technique is characterized by the introduction of irrelevant information that diverts attention from the original argument, distinguishing it from mere statements or expressions that do not engage in debate.

C.14 Appeal to Pity

v1: A technique that 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 soliciting 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.
v2: A technique that evokes feelings of pity, sympathy, compassion, or guilt in the audience to distract it from focusing on evidence, rational analysis, and logical reasoning, thereby leading the audience to accept the speaker's conclusion as truthful based solely on these emotions. It specifically involves claims intended to elicit pity or guilt, rather than merely describing emotional situations, often by presenting vulnerable individuals or distressing circumstances, thereby appealing to the audience's sense of empathy rather than their critical thinking.

C.15 Causal Oversimplification

v1: 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.)
v2: Causal Oversimplification is the technique of assuming a single cause or reason for an issue when there are actually multiple contributing factors. This technique can manifest in claims that suggest a direct causal relationship between two events without considering other influences or causes. It 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).

C.16 False Dilemma or No Choice

v1: Sometimes called the either-or fallacy, a false dilemma is a logical fallacy that presents only two options or sides when there actually are many. One of the alternatives is depicted as a no-go option, and 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).
v2: Sometimes called the either-or fallacy, a false dilemma is a logical fallacy that presents only two options or sides when there actually are many, often framing one option as undesirable or impossible, which leads the audience to believe that the only viable choice is the other option. One of the alternatives is depicted as a no-go option, reinforcing this perception and forcing the audience to choose 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). This technique simplifies complex issues into binary choices, ignoring other possibilities and nuances. It can be identified when the author explicitly limits the options to two, disregarding other possibilities. This technique is characterized by the clear presentation of limited options, where the audience is led to believe that they must choose one of the presented alternatives without considering other possibilities. It often involves framing one option as necessary while dismissing the other as irrelevant or undesirable. A clear indication of this technique is the absence of any mention of alternative options or the framing of a situation as having only two possible outcomes. It often emphasizes the negative consequences of the rejected option to strengthen the perceived necessity of the chosen option and involves framing a situation in such a way that opposing a proposed action is equated with endorsing a negative outcome. Additionally, it implies that any deviation from the presented choices is invalid or unacceptable, further constraining the audience's perception of available options and ignoring other potential solutions or outcomes. Furthermore, this technique implies that if one option is not taken, the other is the only viable choice, disregarding other possibilities and suggesting that the audience must choose between the presented options,

thereby disregarding other viable alternatives. It often creates a sense of urgency or necessity, compelling the audience to choose one of the presented options. This technique can be identified when a statement restricts the options to two, disregarding other possibilities.

C.17 Consequential Oversimplification

v1: 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 dominolike 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).
v2: 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 follows the logical form: if A happens, then B, C, D, ... will happen, often leading to exaggerated or ludicrous negative outcomes while ignoring the likelihood of these events occurring. The core essence behind this fallacy is an assertion of some 'first' event/action leading to a domino-like chain of events that have significant negative effects and consequences, while ignoring and/or understating the likelihood of the sequence of events from the first event leading to the end point. It can also apply when oversimplifying a positive outcome from a complex situation, leading to misleading conclusions. This technique is characterized by a failure to engage with the validity of the original argument and often relies on an exaggerated portrayal of potential negative consequences, as well as a failure to provide evidence for the claimed causal relationships. Additionally, it can be used to promote certain actions by suggesting they will lead to major positive outcomes, with the promise of a significant event in the end, but the key characteristic remains the lack of substantiation for the causal links. This technique typically oversimplifies the likelihood of the events occurring and ignores the complexity and nuances of the situation.

C.18 False Equivalence

v1: A technique that attempts to treat scenarios that are significantly different as if they had equal merit or significance. In particular, an emphasis is being made on one specific shared characteristic between the items of comparison in the argument that is way off in the order of magnitude, oversimplified, or just that important additional factors have been ignored. The introduction of the certain shared characteristics of the scenarios is then used to consider them equal. This technique has the following logical form: A and B share some characteristic X. Therefore, A and B are equal.
v2: A technique that attempts to treat scenarios that are significantly different as if they had equal merit or significance, particularly by emphasizing one specific shared characteristic between the items of comparison in the argument that is oversimplified or ignores important additional factors. The introduction of this shared characteristic is used to argue that the scenarios are equal, typically following the logical form: A and B share some characteristic X. Therefore, A and B are considered equal, despite significant differences in context or implications. This technique is evident when an argument states that A and B share some characteristic X, leading to the conclusion that A and B are equal, even when the contexts of A and B are fundamentally different. It is important to note that the technique is only applicable when a comparison is explicitly made between two distinct scenarios in a way that suggests they are being treated as equivalent. The technique is not present if the comparison does not imply equality or if it critiques without equating. Additionally, this technique can lead to a distorted understanding of the issues at hand by failing to acknowledge the complexities and nuances that differentiate the scenarios.

C.19 Slogans

v1: A brief and striking phrase that may include labeling and stereotyping. Slogans tend to act as emotional appeals.
v2: A brief and striking phrase that may include labeling and stereotyping, often used to create emotional appeals. Slogans are typically memorable and concise, serving as a rallying cry or persuasive statement. They aim to influence public opinion or behavior by simplifying complex ideas into catchy phrases that resonate with the audience, encapsulating a larger message or sentiment in a way that influences public perception. Crafted to resonate emotionally, slogans enhance their impact and recall, making them powerful tools for communication, specifically designed to persuade or influence opinions and attitudes. Slogans are particularly effective in advertising and political campaigns, capturing attention quickly and conveying a clear message. This makes them integral to campaigns that seek to drive action or change perceptions, often used strategically to create a lasting impression and shape brand identity or political narratives. They are typically found in persuasive contexts rather than purely factual statements, aiming to provoke a strong emotional response and influence attitudes and behaviors. Slogans tend to act as emotional appeals, presented as standalone statements in a recognizable and memorable format, further reinforcing their effectiveness in persuasion. Additionally, slogans are designed to encapsulate a message succinctly, emphasizing their role in influencing opinions and attitudes. They often act as emotional appeals, effectively conveying a message in a concise manner.

C.20 Conversation Killer

v1: 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.
v2: This includes words or phrases that discourage critical thought and meaningful discussion about a given topic by presenting it as an undeniable fact, often by oversimplifying complex issues, asserting a false consensus, or denying the existence of disagreement. They dismiss opposing viewpoints as repetitive or unoriginal, thereby shutting down further inquiry. They often manifest as dismissive statements, oversimplifications, or generalizations that shut down further dialogue, and are a form of Loaded Language, often passing as folk wisdom, intended to end an argument and quell cognitive dissonance. Additionally, they may be used strategically to reinforce existing beliefs and discourage any exploration of alternative perspectives, indicating a clear intent to dismiss opposing views and prevent further dialogue. The presence of such language should be evident in the text to classify it as a Conversation Killer. Examples include statements that simplify complex issues, assert a false consensus, label opposing arguments as unoriginal, or use dismissive language that prevents further discussion, which further illustrates their role in stifling discussion. The absence of such language in a text indicates that the Conversation Killer technique is not present.

C.21 Appeal to Time

v1: The argument is centered around the idea that time has come for a particular action. The very timeliness of the idea is part of the argument.
v2: The argument is centered around the idea that time has come for a particular action, explicitly indicating that the current moment is significant for the argument being made and emphasizing the urgency and necessity of acting in the present context due to a specific time-related situation. The very timeliness of the idea is part of the argument, often suggesting that delay could result in missed opportunities or negative consequences. Additionally, it may invoke a sense of immediacy, explicitly stating that immediate action is necessary due to current circumstances, and implying that the current context or situation makes the action particularly relevant or necessary right now. It highlights that the urgency and appropriateness of the timing in relation to the action being proposed are crucial components of the argument, reinforcing the idea that immediate action is essential. The emphasis on timeliness serves as a persuasive element, urging individuals to recognize the importance of acting without delay. This technique can also leverage societal or cultural pressures that prioritize promptness, further enhancing the perceived necessity of immediate action. Furthermore, it may appeal to the audience's emotions by creating a sense of fear or anxiety about the consequences of inaction,

thereby strengthening the call for immediate response or change. Importantly, the argument focuses on the urgency and timeliness of the action itself, rather than merely discussing the consequences of inaction over time, and emphasizes the appropriateness of the idea in the present moment, highlighting the significance of the current moment in relation to the proposed action and its connection to past events. Additionally, it underscores the importance of addressing current issues, reinforcing the notion that the present context demands immediate attention and action, and suggesting that the current moment is critical for the proposed action. The argument should clearly indicate that a specific action is being advocated for at this moment in time, with the timeliness of the idea being a crucial part of the argument, emphasizing that the urgency and timeliness of the idea are critical elements of the overall persuasion.

C.22 Loaded Language

v1: 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.
v2: Loaded Language technique involves the use of specific words and phrases with strong emotional implications (either positive or negative) to influence and convince the audience that an argument is valid or invalid, particularly when such language is used to evoke a strong emotional response or bias in the audience. It is characterized by emotionally charged language that seeks to provoke a strong reaction, aiming to bypass logical reasoning and appeal directly to the audience's feelings. This technique can manipulate perceptions by framing issues in a way that elicits specific emotional reactions, often leading to biased interpretations of the argument presented. Additionally, it can sway opinion by leveraging the emotional weight of language to create a sense of urgency or importance around the issue at hand, often by evoking strong feelings such as shame, pride, anger, sympathy, or admiration, and may involve the use of vivid imagery to enhance the emotional impact of the message. It is important to note that loaded language can also lead to oversimplification of complex issues, as it may reduce nuanced arguments to emotionally charged slogans or catchphrases. Furthermore, loaded language is characterized by the presence of emotionally charged terms that seek to elicit a strong reaction rather than present objective facts, often by evoking strong feelings and moral judgments, and it sways opinion rather than providing a balanced view of the argument. The presence of such language must be evident in the text to determine its use, and this technique is particularly effective when such language is used to provoke strong reactions, specifically designed to influence the audience's perception and emotional response, aiming to sway opinions or feelings. It is also known as Appeal to Argument from Emotive Language. The presence of emotionally charged language rather than neutral or simple expressions is essential for identifying this technique, often characterized by the absence of neutral language. The technique often employs biased language to reinforce emotional appeals, emphasizing the emotional implications of the language used, and it specifically targets the audience's feelings related to the subject matter to enhance its persuasive impact. Additionally, loaded language often involves framing individuals or groups in a particular light, which can further influence the audience's perception and emotional response, and it can evoke fear, anger, or pride to strengthen its persuasive effect.

C.23 Obfuscation, Intentional Vagueness, Confusion

v1: This fallacy uses words that are deliberately not clear, 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 question posed fall under this category too.
v2: This fallacy uses words that are deliberately not clear, 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 question posed fall under this category too. The use of vague terms or phrases that can lead to confusion about their meaning is a key indicator of this technique. Additionally, phrases that lack

specificity and can lead to confusion about their meaning or implications are also considered part of this technique. Furthermore, the use of ambiguous terms or phrases that lack specific meaning can lead to confusion and misinterpretation. Phrases that describe situations without specific details or clarity can also exemplify this technique, further obscuring the intended message and making it difficult for the audience to discern the actual argument being made. Moreover, phrases that create confusion about the severity or nature of a subject can also exemplify this technique, adding to the overall vagueness and misinterpretation of the argument. The use of ambiguous terms or references that lack specificity contributes to confusion and misinterpretation. The use of ambiguous terms or phrases that can be interpreted in various ways contributes to the obfuscation of the argument.

C.24 Exaggeration - Minimisation

v1: 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.
v2: 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. It can also involve using hyperbolic language to emphasize negative actions or consequences while minimizing the perceived importance of accountability. The technique is characterized by clear instances of hyperbole or minimization in the language used, and it can be identified through specific phrases that amplify or diminish the perceived reality of a situation. Additionally, it may involve emphasizing extreme negative situations while neglecting any positive aspects or responses, further skewing the audience's understanding. The technique is evident when there are clear examples of inflated claims or significant downplaying of issues, often leading to a distorted perception of reality.

C.25 Repetition

v1: The speaker uses the same word, phrase, story, or imagery repeatedly with the hope that the repetition will lead to persuade the audience.
v2: The speaker uses the same word, phrase, story, or imagery repeatedly, at least twice, within a context to persuade the audience, emphasizing the importance or urgency of the message. This includes instances where the repetition is clearly aimed at reinforcing a point or argument, and it encompasses cases where the same element is reiterated multiple times throughout the text, particularly in close proximity, to reinforce the message. The repetition must be evident in the text, specifically through clear and noticeable instances of repetition, emphasizing the emotional or thematic significance of the repeated elements. Additionally, this includes instances where the same elements are reiterated in close proximity to further emphasize a point, particularly through the use of specific phrases that highlight key points or failures. The repetition must occur multiple times within the text to effectively contribute to the persuasive impact, specifically through clear and intentional reiteration of key elements, while also highlighting the importance of the repeated elements and their emotional or thematic significance. The definition also emphasizes the act of repeating these elements multiple times within the discourse as a critical aspect of the technique, specifically by emphasizing key points through reiteration, and it highlights the need for the audience to notice the repeated elements, thereby enhancing the perceived urgency of the message. The speaker employs this technique with the hope that the repetition will lead to persuasion, specifically focusing on instances where the same element is reiterated to emphasize a point, and this repetition must be clearly identifiable within the text. Furthermore, this technique also includes instances where the same concept is reiterated in different forms or contexts, allowing for a broader interpretation of the repeated ideas, including cases where the same concept is reiterated without significant variation. This includes instances where the same concept is emphasized through multiple occurrences, reinforcing the overall persuasive effect, specifically by emphasizing key points or themes through their recurrence, and it also includes cases where the repetition is intended

to evoke an emotional response. Additionally, the repetition may be intended to create a rhythm in the message, further enhancing its persuasive quality. This includes instances where the same element is used multiple times within a text or speech, reinforcing the overall message, and it also includes instances where the same element is emphasized multiple times throughout the text, specifically emphasizing the persuasive effect of such repetition. This definition also highlights the importance of direct repetition of the same concept, rather than merely mentioning similar ideas, to strengthen the persuasive impact, specifically through the use of identical or very similar elements in close proximity. If no such repetition is present, the technique is not considered used, as the absence of such repetition indicates that the technique is not applied. This can be identified by the presence of at least one instance of such repetition in the text, specifically with the hope that the repetition will lead to persuading the audience, particularly by emphasizing key terms or concepts that are central to the argument; this can be identified by the presence of identical or similar elements appearing multiple times in the text, particularly within a short span of text, specifically emphasizing the same elements multiple times for effect, and specifically emphasizing the impact of the repeated elements on the audience's perception.