

Analyzing the Effects of Domain-Adaptive Pretraining on Propaganda Technique Detection in a Low-Resource Language

Ieva Rizgelenė, Hannele Seeck

Abstract—This study presents one of the first systematic investigations of domain-adaptive pretraining for propaganda technique detection in Lithuanian. We perform masked language modeling-based domain-adaptive pretraining on a corpus of 63,648 news articles from non-credible outlets. We evaluate its impact using masked-language probing, propaganda technique recognition through sequence tagging and sentence classification, and comparison with a large language model in a zero-shot setting. In addition to quantitative evaluation, we conduct a qualitative analysis based on a case study of annotation behavior. The results show that domain-adaptive pretraining improves propaganda technique detection in Lithuanian, enhances the ranking of contextually meaningful terms, and improves performance for most techniques, with gains varying across tasks.

Impact Statement—Propaganda techniques are key tools used to create manipulative content. Such content tends to be more effective when disseminated in native languages. However, while recent advances have improved propaganda detection in high-resource languages, low-resource languages still face significant limitations due to insufficient NLP resources. This study addresses this gap by investigating the impact of domain-adaptive pretraining on propaganda technique detection in a low-resource language, specifically Lithuanian.

Index Terms—domain-adaptive pretraining, propaganda techniques, low-resource language, transformers, NLP, masked language

I. INTRODUCTION

THE rapid spread of deceptive content on social media has intensified concerns about the societal impact of online propaganda. Unlike misinformation, which primarily involves factual inaccuracy, propaganda often relies on rhetorical manipulation that may incorporate factual elements while still misleading audiences through specific persuasive techniques. Consequently, simple “true vs. false” fact-check warnings may be insufficient to address the nuanced nature of propagandistic content. Recent experimental research demonstrates that labeling posts as propaganda and explicitly highlighting the rhetorical techniques employed can significantly reduce users’ intentions to share such content online [1]. These

This work was supported in part by NordForsk and the Research Council of Lithuania (LMTLT) under Agreement No. S-NORDFORSK-26-5 for the BIO-NORD project (Project No. 233158), and in part by the Helsingin Sanomat Foundation.

Ieva Rizgelenė is with the Institute of Data Science and Digital Technologies, Vilnius University, Vilnius, Lithuania (e-mail: ieva.rizgeliene@mif.vu.lt).

Hannele Seeck is with Department of Social Sciences, LUT University, Finland (e-mail: hannele.seeck@lut.fi) and with Department of Media and Communications, The London School of Economics and Political Science, UK (e-mail: h.seeck@lse.ac.uk)

findings suggest that enabling users to recognize propaganda techniques may help reduce the dissemination of non-credible content.

Interest in the automated detection of propaganda techniques has grown substantially in recent years, as reflected in a series of shared tasks dedicated to propaganda and persuasion technique detection [3], [4], [5], [6], [7], [8]. Despite this progress, significant challenges remain, particularly for low-resource languages. A central limitation is the scarcity of modeling resources, including annotated propaganda corpora and pretrained language models tailored to these languages. In addition, pretrained transformers are typically trained on general-domain corpora (e.g., Wikipedia) and may therefore provide incomplete coverage of specialized domains such as non-credible or propagandistic discourse. Continued pretraining on in-domain data, commonly referred to as domain-adaptive pretraining (DAPT), has been shown to substantially improve downstream performance across tasks [9]. Empirical studies further indicate that domain-adapted models can enhance propaganda detection performance in high-resource languages [11], [12].

However, while DAPT has demonstrated promising results in high-resource settings, its effectiveness for fine-grained propaganda technique detection in low-resource languages has not been systematically examined. This gap has broader implications beyond model performance. Low-resource languages are often underrepresented in cross-lingual NLP systems, which can lead to uneven performance in propaganda detection pipelines across languages. As a result, propaganda detection systems may be less effective in linguistic and cultural contexts where they are critically needed. In particular, it remains unclear how domain adaptation affects not only overall performance but also the model’s internal contextual representations and prediction patterns.

In this study, we investigate the role of DAPT for propaganda technique detection in Lithuanian, a low-resource language. Our work is guided by three main research questions:

RQ1: Does DAPT improve propaganda technique detection performance in a low-resource language?

RQ2: How does DAPT alter the model’s internal contextual representations of technique-specific lexical and semantic cues?

RQ3: How does domain adaptation influence annotation behavior in fine-tuned transformer models?

In this work, we use the term *annotation behavior* to refer to how the model assigns propaganda technique labels to tokens,

spans, and sentences, i.e., how it behaves as an automatic annotator when marking propagandistic content.

To address these questions, we further pretrain the Lithuanian transformer LT-MLKM-modernBERT on 63,648 domain-specific news articles from non-credible sources using masked language modeling. Prior to task-specific fine-tuning, we analyze representational differences between the original pretrained model and its domain-adapted counterpart through masked word prediction experiments across nine propaganda techniques. We then fine-tune both models for token-level propaganda technique tagging and sentence-level classification and conduct both quantitative and qualitative evaluations. The quantitative analysis includes MLM-based ranking analysis and performance evaluation for the sequence tagging and sentence classification tasks, while the qualitative analysis examines annotation behavior through a case study.

II. RELATED WORK

Fine-grained automated propaganda technique detection is commonly traced to the NLP4IF-2019 shared task, which framed the problem as sentence-level classification and fragment-level span identification for 18 propaganda techniques [2]. SemEval-2020 Task 11 extended this line of work to span identification and technique classification in news articles [3]. Subsequent SemEval tasks broadened the scope to persuasion technique detection in multimodal content (text and images) in 2021 [4], multilingual news analysis in 2023 [5], and multilingual meme-based persuasion detection in 2024 [6]. Beyond SemEval, related initiatives have addressed non-English and multilingual settings, including Arabic shared tasks on technique classification and persuasion detection in tweets [7], [17] and the DIPROMATS shared tasks on propaganda identification, characterization, technique classification, and narrative detection in English and Spanish [18], [25].

Recent studies have further explored propaganda and misinformation detection using transformer-based models [22], [23]. In addition, graph-based approaches have been proposed to integrate syntactic, semantic, and sequential relations across words for propaganda detection [27], while meta-learning methods combined with pretrained language models have also been applied to propaganda technique detection [28].

Collectively, these efforts have established important datasets, taxonomies, and evaluation benchmarks for propaganda and persuasion identification. However, the primary goal of these studies has been comparative evaluation across datasets, languages, and model architectures. Even in multilingual settings, the emphasis is largely on benchmarking cross-lingual generalization rather than on systematically studying domain adaptation strategies and their effects in low-resource scenarios. This distinction is important for the present work: multilingual shared tasks broaden language coverage, but they do not analyze the contribution of DAPT or examine how it alters model behavior.

DAPT itself was established as a general adaptation strategy [9], which showed that continued pretraining on in-domain unlabeled text can improve downstream performance. In propaganda detection, the clearest direct application in our

reviewed literature is the SemEval-2020 system of [11], which further pretrained BERT on propagandistic and hyperpartisan news before fine-tuning it for propaganda detection, reporting improvements for several techniques. While this result suggests that DAPT can benefit propaganda-related tasks, it was obtained in a high-resource-language shared-task setting, specifically English, and was not presented as a systematic study of DAPT for fine-grained propaganda technique detection.

Evidence from related tasks further supports the usefulness of DAPT. Continued pretraining has improved media bias detection [24], political ideology prediction [14], and hate speech detection, including BERT models adapted to Reddit data from banned communities [10]. DAPT has also shown promise in low-resource settings: domain-adapted transformer models improved Urdu fake news detection across multiple datasets [19], and DAPT applied to Bangla BERT yielded gains in multiclass hate speech detection [15]. However, these studies address document- or instance-level classification tasks in related domains rather than fine-grained propaganda technique detection, and they do not examine the effects of DAPT on internal model representations or annotation behavior.

Research Gap and Motivation

Taken together, prior work falls into two main strands: (i) shared tasks that advance multilingual benchmarking for propaganda and persuasion analysis, and (ii) DAPT studies conducted either in high-resource propaganda settings or in related low-resource classification tasks. To the best of our knowledge, no prior study has systematically evaluated DAPT for fine-grained propaganda technique detection in a low-resource language while also analyzing its impact on model representations and annotation behavior. This gap is particularly pronounced in the Nordic-Baltic context. For Lithuanian, such work has only recently become feasible with the introduction of the first annotated propaganda corpus [16], whereas for other Nordic-Baltic languages the lack of comparable datasets continues to constrain progress. Our study addresses this gap and provides evidence on how DAPT can support fine-grained propaganda analysis in low-resource language settings.

III. METHODS

A. Domain-Adaptive Pretraining (DAPT)

DAPT is a continued pretraining strategy in which a pretrained transformer-based language model is further trained on an additional corpus of unlabeled texts drawn from a specific target domain. During DAPT, the model continues to optimize the same self-supervised objective used in the original pretraining, typically Masked Language Modeling (MLM), while shifting the training distribution toward domain-relevant data.

Formally, given pretrained parameters θ_0 obtained from general-domain data, DAPT continues optimization from θ_0 by minimizing the same self-supervised loss over domain-specific texts $\mathcal{D}_{\text{domain}}$:

$$\theta_{\text{DAPT}} = \arg \min_{\theta} \mathbb{E}_{x \sim \mathcal{D}_{\text{domain}}} [\mathcal{L}_{\text{MLM}}(\theta)],$$

where $\mathcal{L}_{\text{MLM}}$ denotes the masked language modeling objective. The optimization is initialized with $\theta = \theta_0$.

B. Masked Language Modeling (MLM)

MLM is a self-supervised pretraining objective commonly used in encoder-based transformer models. The objective enables the model to learn contextualized representations of language without requiring labeled data.

Given an input token sequence

$$x = (x_1, x_2, \dots, x_n),$$

a subset of tokens $M \subset \{1, \dots, n\}$ is randomly selected and replaced with a special [MASK] token. The model is then trained to predict the original tokens at the masked positions using the surrounding context.

Formally, the MLM objective minimizes the negative log-likelihood of the masked tokens:

$$\mathcal{L}_{\text{MLM}}(\theta) = - \sum_{i \in M} \log P_{\theta}(x_i | x_{\setminus M}),$$

where $x_{\setminus M}$ denotes the input sequence with masked positions, and $P_{\theta}(x_i | x_{\setminus M})$ represents the probability assigned by the model to the correct token x_i given its context.

C. LT-MLKM-modernBERT Model (BASE Model)

In this study, we employ *LT-MLKM-modernBERT* [26] as the base language model for both DAPT and the technique-tagging task. LT-MLKM-modernBERT is a monolingual, encoder-only transformer based on the *ModernBERT-base* architecture and specifically pretrained for the Lithuanian language. Pretraining was conducted on approximately 1.87 billion words (around 49 billion tokens) collected from diverse Lithuanian sources, including news, legal, academic, and public sector texts. LT-MLKM-modernBERT comprises 22 Transformer encoder layers with 12 attention heads and a hidden representation size of 768 dimensions, resulting in approximately 149 million parameters. The model employs a custom Lithuanian tokenizer with a vocabulary of 64,000 tokens and supports a maximum input sequence length of 8,192 tokens.

IV. DATA

A. HALT-PROP corpus

For the propaganda technique tagging and sentence classification tasks, we used the HALT-PROP dataset [16], the first human-annotated Lithuanian corpus for propaganda narratives and techniques. The corpus contains 1,000 annotated articles for these ten propaganda techniques:

- 1) Emotional Expression – Uses emotionally charged wording to provoke feelings (fear, anger, pride).
- 2) Whataboutism / Red Herring / Straw Man – Diverts attention by deflecting, introducing irrelevant points, or misrepresenting the opponent.
- 3) Simplification – Reduces complex issues to overly simple causes, slogans, or false choices.

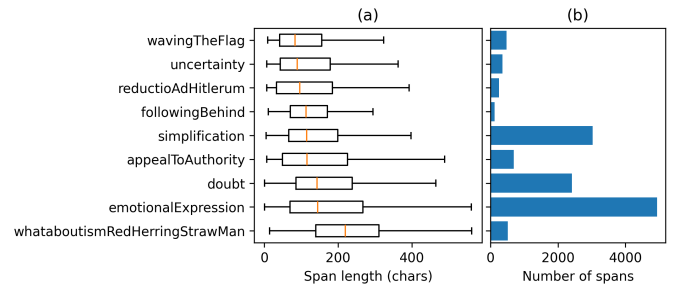


Fig. 1. HALT-PROP corpus statistics: (a) Span length by technique; (b) Span frequency by technique.

- 4) Intentional Vagueness (Uncertainty) – Uses unclear or ambiguous wording to avoid specifics and accountability.
- 5) Appeal to Authority – Justifies a claim by citing authority instead of evidence.
- 6) Flag-Waving – Promotes a stance using patriotism or national pride.
- 7) Bandwagon – Encourages agreement by claiming “everyone supports it.”
- 8) Doubt / Smears – Undermines credibility through suspicion, insinuations, or personal attacks.
- 9) Reduction ad Hitlerum/Stalinum – Discredits by associating with Hitler/Stalin or similar notorious figures.
- 10) Repetition – Repeats a message to make it seem more true or convincing.

However, in this study, we excluded the *Repetition* technique, because it has a different usage pattern in articles. Unlike other techniques, it is not defined only by individual annotated spans, but also by relations between multiple occurrences of the same message, word, or phrase across the document. Therefore, detecting *Repetition* requires not only sequence tagging or classification, but also document-level linking of repeated expressions, which goes beyond the scope of this work. For this reason, we focus on the remaining nine techniques, which can be addressed using a standard approach.

Figure 1 presents the distribution of propaganda technique spans in the HALT-PROP corpus after excluding the *Repetition* technique. As shown, span lengths vary substantially across techniques. *Whataboutism/Red Herring/Straw Man* exhibits the longest spans on average, indicating that it is typically expressed through larger argumentative fragments. In contrast, techniques such as *Flag-Waving*, *Uncertainty*, and *Following Behind* are characterized by shorter spans. Figure 1(b) shows a strong class imbalance in the dataset: *Emotional Expression*, *Simplification*, and *Doubt* are the most frequent techniques, while others occur considerably less often.

B. DAPT Data

For DAPT, we used 63,648 Lithuanian news articles crawled from six non-credible news outlets corresponding to the sources in the HALT-PROP corpus. We also excluded 1,000 articles that had been annotated for the HALT-PROP corpus. The collected articles cover the period 2017–2024. Fig. 2 presents the distribution of the corpus by source and publication year.

TABLE I
SOURCE CATEGORIES WITH ARTICLE COUNTS (n) AND AVERAGE ARTICLE LENGTH (IN CHARACTERS).

Source	Categories
1	Opinions ($n = 811$, avg.=6144); Lithuania ($n = 738$, avg.=3587); Politics ($n = 494$, avg.=3151); Leisure ($n = 139$, avg.=4534); World ($n = 125$, avg.=2136); Law Enforcement ($n = 120$, avg.=4395); Health ($n = 63$, avg.=5443); Business ($n = 48$, avg.=3673)
2	World ($n = 5092$, avg.=2093); Technology ($n = 721$, avg.=2285); Health ($n = 418$, avg.=3962); Lithuania ($n = 406$, avg.=3007); Family ($n = 229$, avg.=3956)
3	News ($n = 6191$, avg.=2361); Politics ($n = 3870$, avg.=2155); World ($n = 1044$, avg.=1588); Traditions ($n = 986$, avg.=2293); Health ($n = 921$, avg.=2702); Samogitia ($n = 627$, avg.=2347); Interesting ($n = 595$, avg.=1900); Psychology ($n = 587$, avg.=2759)
4	Politics ($n = 10899$, avg.=2518); Opinions ($n = 3319$, avg.=3883); War and Peace ($n = 1836$, avg.=2461); Foreign Affairs ($n = 1796$, avg.=1274); GGN Publicism ($n = 410$, avg.=1092); Anti-Mafia ($n = 228$, avg.=2741); Topical Call ($n = 173$, avg.=1232); Pedophilia ($n = 173$, avg.=2854)
5	Current Affairs ($n = 4104$, avg.=3983); Lithuania ($n = 3466$, avg.=2893); World ($n = 3027$, avg.=2805); Lifestyle ($n = 1323$, avg.=2792); Opinions ($n = 564$, avg.=2852); Wellness ($n = 461$, avg.=4184); Leisure ($n = 459$, avg.=4439); Science ($n = 370$, avg.=3341)
6	Unknown ($n = 19397$, avg.=3588)

The dataset is dominated by three major sources, while the remaining outlets contribute substantially fewer articles. Most articles were published between 2018 and 2024, with only a small number dated 2017 and a slightly lower volume observed in 2024. This temporal pattern reflects the activity of these outlets, which became more active after 2018, while the decline in 2024 is consistent with several sources reducing operations or shutting down due to legal proceedings related to the dissemination of malicious propaganda.

Despite being categorized as propagandistic or non-credible, a large portion of the collected content consists of general news rather than explicitly propagandistic material. Table I summarizes the distribution of articles across categories, reporting both the number of articles and the average article length. Category-level analysis was not available for Source 6, as it did not include category metadata. For the remaining sources, we observe a wide range of categories, indicating that the corpus contains a mixture of general news and potentially propagandistic content. Categories such as Health, Lifestyle, and Leisure suggest the presence of general news, while others, including Anti-Mafia and Current Affairs, are more likely associated with ideologically driven or propagandistic content.

Article length varies considerably across categories, ranging from relatively short texts (approximately 1,000–1,500 characters) to longer articles exceeding 6,000 characters. In general, categories associated with commentary or opinion tend to be longer (e.g., Opinions in Source 1, avg. 6144), whereas some categories contain notably shorter texts, such as Foreign Affairs (avg. 1274), Topical Call (avg. 1232), and GGN Publicism (avg. 1092). This suggests a mix of concise reports and more elaborate articles.

In addition, the distribution of articles across categories is

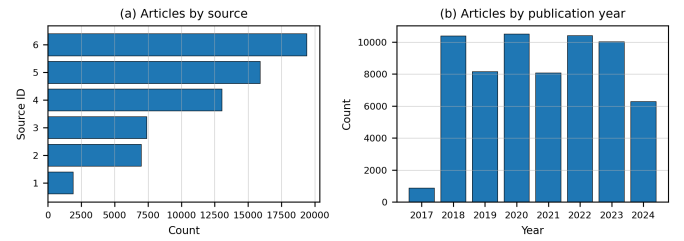


Fig. 2. DAPT corpus statistics: (a) articles by source; (b) articles by publication year.

highly imbalanced, with a large proportion concentrated in a small number of dominant categories such as News, Politics, and World, while many other categories contain substantially fewer articles.

Overall, these observations indicate that the corpus is heterogeneous in both content and style, combining short- and long-form texts across diverse thematic domains. Consequently, DAPT on this corpus exposes the model not only to propaganda-related discourse but also to a broad range of general Lithuanian news language. Finally, we emphasize that this corpus was not included in the original pretraining data of LT-MLKM-modernBERT and therefore provides additional domain-specific Lithuanian context during adaptation.

V. EXPERIMENTAL SETUP

As our main goal is to assess the impact of DAPT on propaganda technique detection in a low-resource language, we not only fine-tune the models for the technique tagging or classification task, but also conduct an additional analysis using the MLM objective. This analysis allows us to compare how the BASE and DAPT models recognize technique-specific terms from contextual information.

Our experimental setup consists of three main stages: (1) applying DAPT to the BASE model to obtain the DAPT model, (2) an MLM-based analysis comparing the BASE and DAPT models, and (3) fine-tuning the models for propaganda technique detection.

A. First Stage: DAPT

First, a small validation split of 2% was created from the DAPT corpus to monitor training progress. The pretrained tokenizer associated with *LT-MLKM-modernBERT* was applied without truncation or padding during the initial tokenization step. All tokenized sequences were then concatenated and divided into fixed-length blocks of 512 tokens.

Training followed the standard MLM objective. Dynamic masking was applied with a masking probability of 15%, meaning that in each training batch, 15% of tokens were randomly selected and replaced according to the standard MLM strategy. The model was trained to predict the original masked tokens based on their surrounding context.

Training was conducted for three epochs using the AdamW optimizer with a learning rate of 5×10^{-6} , weight decay of 0.01, and 500 warm-up steps. The batch size was set to 8 for both training and validation. The checkpoint with the lowest validation loss was selected as the final model.

B. Second Stage: MLM-Based Analysis of BASE vs. DAPT Models

In this stage, we use the original pretrained *LT-MLKM-modernBERT* as the BASE model and its domain-adapted version obtained after DAPT (the model resulting from the first stage of training).

For each propaganda technique, we first extracted the most frequent lexical items from the HALT-PROP corpus, focusing on four parts of speech: adjectives, nouns, proper nouns, and verbs. These categories were selected because they carry the most semantic and contextual information, capturing key entities, actions, and descriptive attributes, while function words were excluded due to their limited discriminative value across techniques. From the annotated propaganda spans, we then compiled a set of distinct lemmas associated with these techniques.

For the MLM analysis, we further restricted this set to lemmas that are represented as single tokens by the *LT-MLKM-modernBERT* tokenizer. Additionally, only those lemmas whose observed surface forms in the corpus were also tokenized as single tokens were retained. This constraint was introduced to ensure methodological consistency and interpretability, as multi-token words would require masking multiple subword units and introduce additional complexity.

For each selected lemma and its surface forms, we extracted all corresponding annotated propaganda spans for each technique. In these contexts, we applied a masking procedure by replacing the target term with a single [MASK] token. The resulting masked fragment was then provided as input to both the BASE and DAPT models.

For the masked position, each model produces a probability distribution over the entire vocabulary. Based on this distribution, candidate tokens are ranked in descending order of their predicted probability, with the highest-probability token assigned rank 1. We evaluate model performance by measuring the rank position assigned to the correct (gold) token.

C. Third Stage: Propaganda technique detection

For this task, we use the HALT-PROP corpus. First, we separate a test set that is used exclusively for final evaluation. Specifically, 105 articles are selected using stratified sampling by propaganda technique. Instead of formulating the task as a multiclass classification problem, we model each propaganda technique as a separate binary task. This approach better reflects the multi-label nature of the data, as a single textual fragment may express multiple techniques simultaneously.

To provide a more comprehensive evaluation, we explore two modeling approaches for propaganda technique detection: sequence tagging and sentence classification. In the sequence tagging setting, the task is formulated as the detection of exact propaganda technique spans through token-level classification. Each token is classified with respect to a specific target technique. We use a binary tagging scheme, where tokens belonging to annotated spans of the target technique are labeled as positive (1), while all other tokens are labeled as negative (0). Since each technique is modeled separately, the fine-tuning process is repeated independently for each

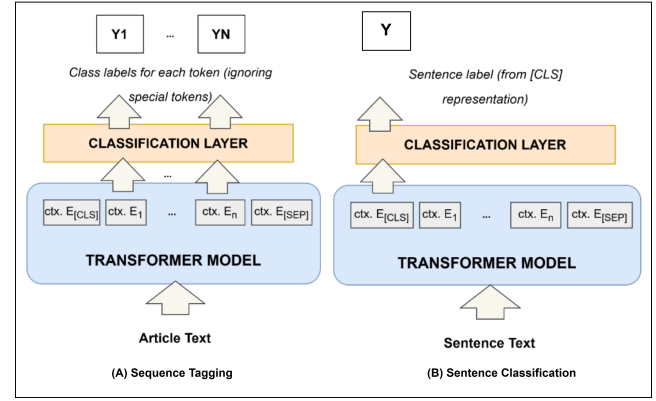


Fig. 3. Task-specific classification using Transformer encoder representations.

propaganda technique, using only the spans annotated for that specific technique as positive examples.

In the sentence classification setting, the text is first split into sentences. Each sentence is then assigned a binary label for each propaganda technique. If a sentence contains at least one annotated span of the target technique, it is labeled as positive (1); otherwise, it is labeled as negative (0). As in the sequence tagging setting, each propaganda technique is modeled separately as a binary classification task.

From the transformer architecture perspective, both approaches use the same input encoding process. The input text is tokenized, encoded, and passed through the transformer encoder layers to obtain contextualized token representations. The main difference lies in the classification head. In sequence tagging, classification is performed independently for each token position, excluding special tokens such as [CLS] and [SEP]. In sentence classification, only the contextualized representation of the [CLS] token is used, as it serves as the sentence-level representation. Figure 3 illustrates this distinction.

For both task formulations, we fine-tune the BASE and DAPT models independently for each propaganda technique. That is, for each technique, we train a separate binary model for sequence tagging and a separate binary model for sentence classification. In each case, the target technique is treated as the positive class, and all other tokens or sentences are treated as negative examples. As mentioned above, the test set is separated from the full corpus and used only during the final evaluation stage. The remaining data is used for training and validation. For each technique, 15% of the remaining data is reserved for validation in order to monitor model performance. The split is performed at the article level to ensure that text chunks from the same article do not appear in both the training and validation sets, thereby preventing data leakage.

The models are trained using AdamW with a learning rate of 5×10^{-6} and weighted cross-entropy loss to address class imbalance. Training is conducted for up to 20 epochs with early stopping based on validation F1 score, using a patience value of 10. The best model is selected according to validation F1.

Performance is evaluated separately for each task. For sequence tagging, token predictions are converted into character-

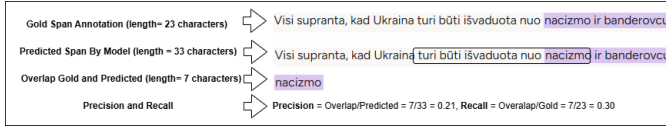


Fig. 4. Example illustrating the calculation of precision and recall for span identification based on character-level overlap between predicted and gold spans.

level spans, and precision, recall, and F1 score are computed based on character overlap between predicted and gold-standard spans. For sentence classification, we report accuracy and Macro-F1. Macro-F1 is used as the main metric because, unlike Micro-F1, it gives equal weight to both classes and is therefore more suitable for highly imbalanced data. The evaluation metrics used in this study are described below.

Overlap-based span F1 score

Let d denote a document, $T_d = \{t_1, \dots, t_n\}$ the set of gold technique spans in d , and $S_d = \{s_1, \dots, s_m\}$ the set of predicted spans. Each span is represented as a contiguous sequence of character indices. Predicted spans are allowed to overlap and are merged prior to evaluation.

Precision and recall are computed by measuring character-level overlap between predicted and gold spans:

$$P = \frac{1}{|S|} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T_d} \frac{|s \cap t|}{|s|}$$

$$R = \frac{1}{|T|} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T_d} \frac{|s \cap t|}{|t|}$$

Precision measures the proportion of predicted characters that are correctly assigned to a technique, while recall measures the proportion of gold characters that are successfully recovered by the model. Figure 4 illustrates an example showing how precision and recall are calculated, providing a clearer understanding of what the metric measures. The final evaluation score is the harmonic mean of precision and recall:

$$F_1 = \frac{2 \cdot P \cdot R}{P + R}.$$

Macro-F1 for binary classification

For the sentence-level classification task, we use the *macro-F1* score, defined as the unweighted average of the F1 scores for the positive and negative classes. This metric assigns equal importance to each class and is therefore more robust than accuracy in the presence of class imbalance.

Let TP, FP, FN, TN denote the counts in the confusion matrix with respect to the positive class ($y=1$). The corresponding F1 scores are:

$$F1_+ = \frac{2TP}{2TP + FP + FN}, \quad F1_- = \frac{2TN}{2TN + FP + FN}.$$

The macro-F1 score is computed as:

$$F1_{\text{macro}} = \frac{1}{2} (F1_+ + F1_-).$$

Additionally, we report *accuracy* as a supplementary metric for general comparison. In the binary setting, accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

Comparison with ChatGPT

For comparison, we evaluate GPT-5.2 as an annotation model using prompt-based inference in *agent mode*. In this setting, the model operates as an autonomous annotation agent that processes input text according to predefined instructions.

To align with our supervised setup, we construct prompts separately for each propaganda technique and for each task formulation. In the zero-shot setting, the prompt includes only a definition of the target technique, without additional examples.

The model is instructed to: (i) analyze the input text independently, (ii) assign labels or identify spans corresponding to the specified technique, and (iii) return the output in a structured JSON format.

For the sequence tagging task, the prompt is formulated to identify spans corresponding to a given technique. An example prompt for the *Doubt* technique is shown below:

Span annotation prompt (Doubt technique):

You are an expert annotator of propaganda techniques.

Analyze the article and identify all text spans that express the technique listed below.

A span is a continuous sequence of words that reflects the technique.

Only analyze the article itself. Do NOT infer meaning beyond the given text. Do NOT invent spans. Do NOT output explanations. Return JSON only.

TECHNIQUE Doubt: Undermines credibility by casting suspicion or attacking character without presenting concrete evidence.

OUTPUT FORMAT

"articleid": articleid, "spans": ["text": "span text", "start": "char start", "end": "char end"]

Rules: - Return all spans expressing the technique - If no spans are present, return an empty list: "spans": [] - Spans must be continuous (no disjoint tokens) - Do not include surrounding neutral text outside the span - Use exact character positions from the text - Output valid JSON only

For the sentence classification task, the prompt is formulated to assign a binary label to each sentence. An example prompt for the *Doubt* technique is shown below:

Sentence classification prompt (Doubt technique):

You are an expert annotator of propaganda techniques.

Analyze each sentence and assign a binary label (0 or 1) for the technique listed below. Label = 1 if the technique is present. Label = 0 if the technique is not present.

Only analyze the sentence itself. Do NOT infer meaning beyond the given sentence. Do NOT invent labels. Do NOT output explanations.

Return JSON only.

TECHNIQUE Doubt: Undermines credibility by casting suspicion or attacking character without presenting concrete evidence.

OUTPUT FORMAT

"articleid": articleid, "sentid": sentid, "sentence": "sentence text", "label": "Doubt": 0 or 1

Rules: - Labels must be 0 or 1 - Do not rename labels - Output valid JSON only

VI. RESULTS

A. DAPT fine-tuned model

The DAPT fine-tuning was conducted on a workstation equipped with an NVIDIA GeForce RTX 4070 Ti SUPER

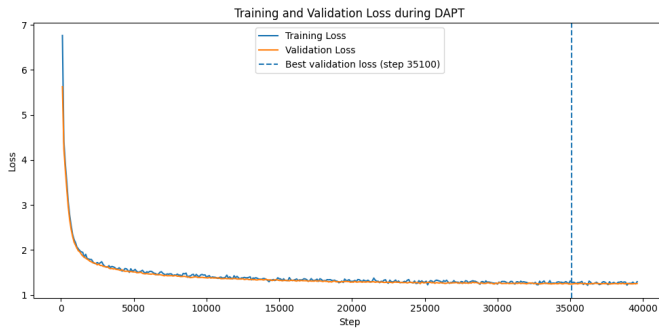


Fig. 5. Training and validation loss during DAPT

GPU (approximately 17 GB VRAM) running CUDA version 12.1. The training process was carried out for 3 epochs, corresponding to a total of 39,612 optimization steps, and required approximately 3 hours and 45 minutes to complete. Evaluation was performed at regular step intervals, and the checkpoint with the lowest validation loss was automatically selected as the final model.

The training dynamics during DAPT show stable convergence (see Figure 5). At the beginning, both training and validation loss decrease sharply (from approximately 6–7 down to approximately 2), which indicates that the model quickly adapted to the domain-specific distribution of the corpus. After this rapid initial phase, the loss continues to decline more gradually, stabilizing around approximately 1.25–1.35. The best validation loss is reached around step 35,100, after which improvements plateau, indicating convergence.

B. MLM analysis

The top adjectives, nouns, proper nouns, and verbs for each technique are shown in Table II. Across all techniques, this yielded 35 distinct lemmas: *naujas*, *politinis*, *didis*, *panašus*, *svarbus*, *tautinis*, *nacistinis*, *pasaulinis*, *valstybinis*, *žmogus*, *valstybė*, *valdžia*, *pasaulis*, *laikas*, *seimas*, *tauta*, *nacis*, *režimas*, *metai*, *prezidentas*, *šalis*, *šeima*, *vaikas*, *Lietuva*, *Rusija*, *Ukraina*, *Europa*, *galėti*, *turėti*, *nebūti*, *kovoti*, *vadinti*, *būti*, *sakyti*, *reikalauti*.

For the ranking analysis, we restricted the set to lemmas represented as a single token by the *LT-MLKM-modernBERT* tokenizer. We further considered only those lemmas whose attested surface forms in the annotated propaganda spans were also tokenized as single tokens. The final set of analyzed lemmas was: *būti* (to be, 38 forms), *Europa* (Europe, 2), *galėti* (can / be able to, 22), *kovoti* (to fight, 8), *laikas* (time, 7), *pasaulis* (world, 6), *prezidentas* (president, 5), *reikalauti* (to demand / require, 8), *sakyti* (to say, 20), *tauta* (nation, 6), *tautinis* (national, 10), *turėti* (to have, 29), *vadinti* (to call / to name, 8), *vaikas* (child, 8), *valstybė* (state, 9), *šalis* (country, 7), and *žmogus* (person, 7). The number in parentheses indicates how many distinct surface forms of the corresponding lemma, observed in the annotated propaganda spans of the corpus, were also single-token forms.

Figure 6 shows the mean rank difference between the DAPT and BASE models for each lemma–technique pair. A negative difference indicates that the DAPT model, on average, assigns

TABLE II
TOP 3 LEMMAS PER POS AND PROPAGANDA TECHNIQUE.

Technique	POS	Lemma 1	Lemma 2	Lemma 3
Emotional Expression	ADJ	naujas	politinis	didis
	NOUN	žmogus	valstybė	valdžia
	PROP	Lietuva	Rusija	Ukraina
	VERB	galėti	turėti	nebūti
Simplification	ADJ	naujas	didis	politinis
	NOUN	žmogus	valstybė	valdžia
	PROP	Lietuva	Rusija	Ukraina
	VERB	galėti	turėti	nebūti
Uncertainty	ADJ	naujas	politinis	panašus
	NOUN	žmogus	pasaulis	laikas
	PROP	Lietuva	Rusija	Ukraina
	VERB	galėti	nebūti	turėti
Doubt	ADJ	politinis	didis	naujas
	NOUN	žmogus	valstybė	seimas
	PROP	Lietuva	Rusija	Europa
	VERB	galėti	turėti	nebūti
Waving the Flag	ADJ	svarbus	tautinis	politinis
	NOUN	valstybė	tauta	žmogus
	PROP	Lietuva	Europa	Rusija
	VERB	turėti	galėti	nebūti
Reductio ad Hitlerum	ADJ	nacistinis	naujas	pasaulinis
	NOUN	nacis	žmogus	režimas
	PROP	Lietuva	Ukraina	Europa
	VERB	kovoti	vadinti	būti
Appeal to Authority	ADJ	didis	naujas	svarbus
	NOUN	žmogus	metai	prezidentas
	PROP	Lietuva	Ukraina	Rusija
	VERB	sakyti	turėti	galėti
Whataboutism / RH / SM	ADJ	didis	naujas	svarbus
	NOUN	žmogus	valstybė	šalis
	PROP	Lietuva	Rusija	Ukraina
	VERB	galėti	turėti	nebūti
Following Behind	ADJ	politinis	svarbus	valstybinis
	NOUN	žmogus	šeima	vaikas
	PROP	Lietuva	Rusija	Europa
	VERB	turėti	reikalauti	galėti

a better rank to the gold term, i.e., a lower position closer to rank 1. A positive difference indicates that the BASE model performs better. The numbers in parentheses denote the number of annotated spans for each technique that contain the selected lemma form. Figure 7 presents the win rates of the DAPT and BASE models. Specifically, it shows the percentage of cases in which each model assigns a lower rank to the gold term, as well as the proportion of cases in which both models assign the same rank.

Overall, the results indicate that DAPT generally improves the ranking of gold terms compared with the BASE model, as negative rank differences are observed in a larger number of lemma–technique combinations (see Figure 6). However, the effect of DAPT is not uniform. Its impact varies across propaganda techniques, parts of speech, and individual lemma–technique combinations.

At the technique level, DAPT tends to provide the strongest improvements for techniques such as *Waving the Flag*, *Appeal to Authority*, *Uncertainty*, and *Emotional Expression*. These techniques often involve contextual framing, rhetorical cues, and emotionally or ideologically loaded language, although such features are not exclusive to them. Nevertheless, the win-rate analysis shows a more balanced pattern: although DAPT

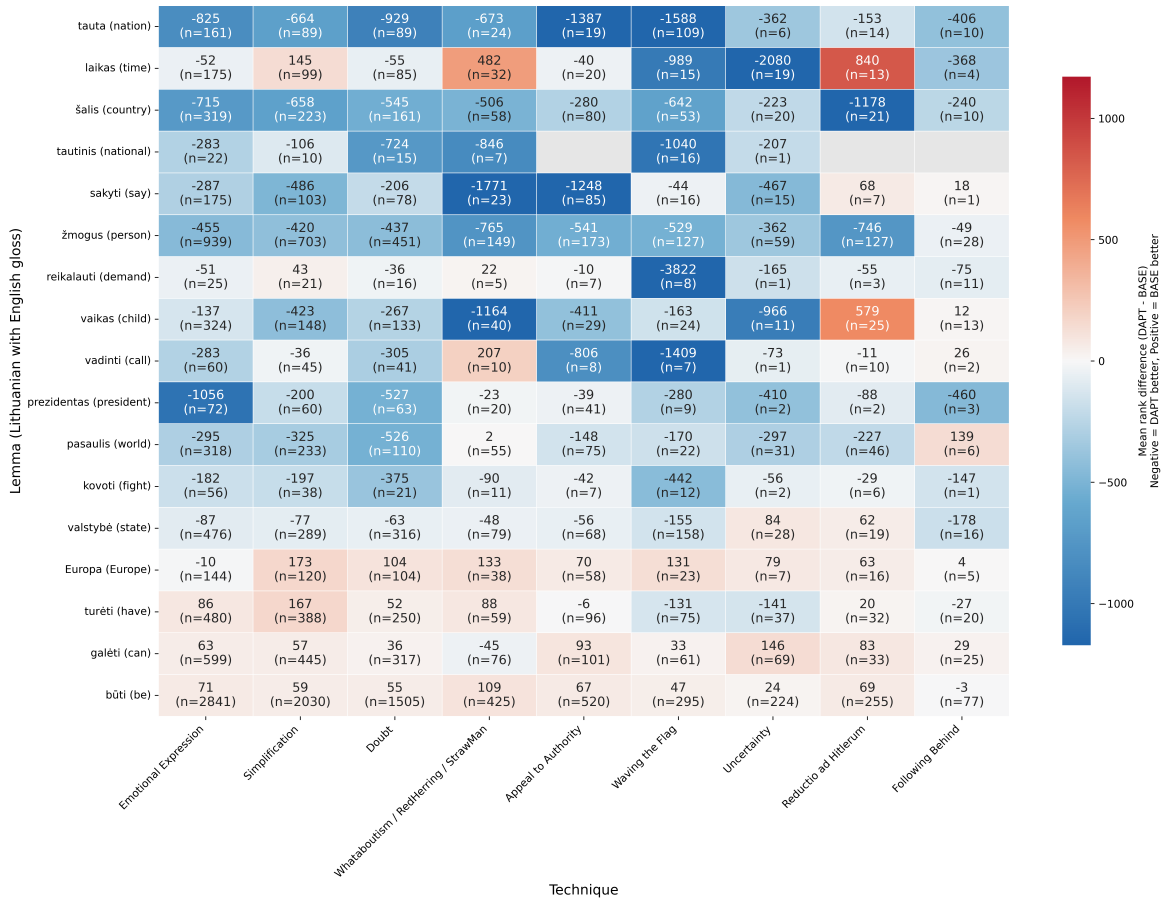


Fig. 6. Mean rank differences between DAPT and BASE models for selected lemmas across propaganda techniques. Negative values indicate higher ranking by DAPT, while positive values indicate better ranking by BASE; cell annotations show frequency (n) of lemma occurrences.

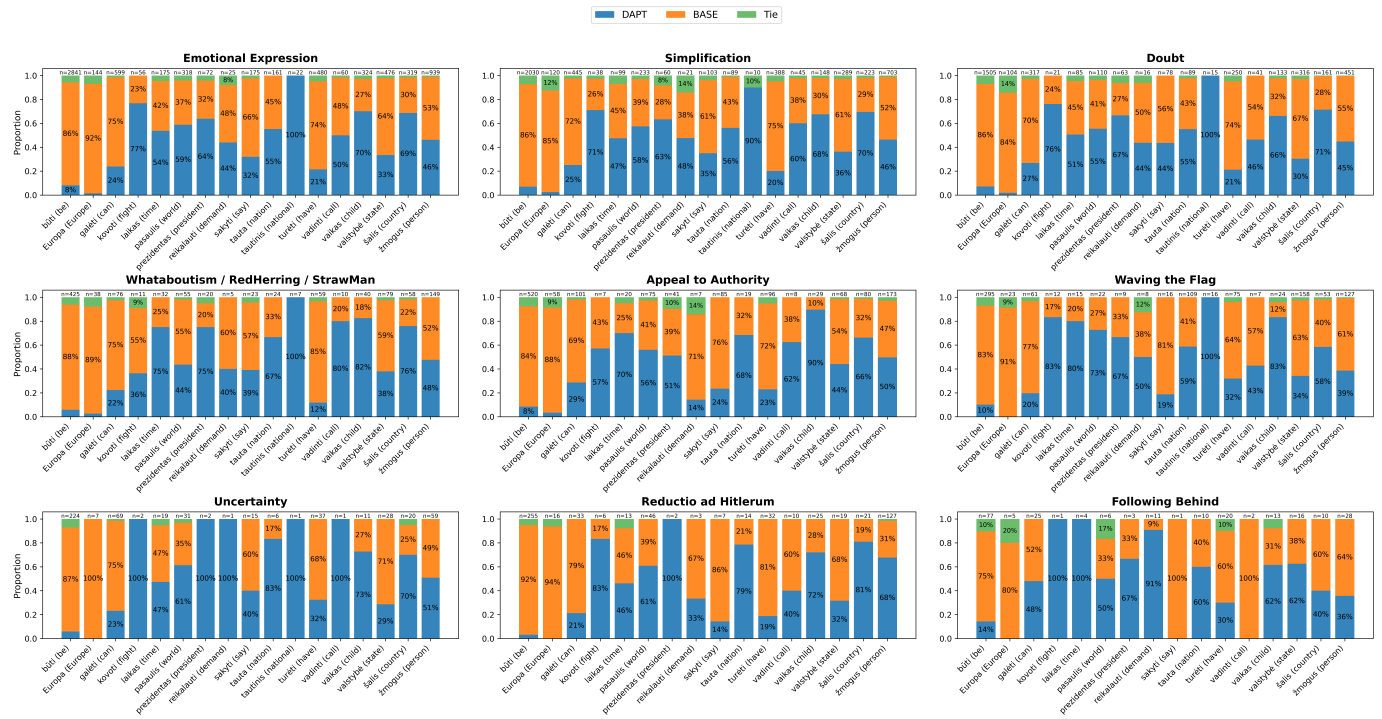


Fig. 7. Comparison of DAPT and BASE win rates across propaganda techniques for top lemmas. Bars show the proportion of cases where each model assigns a better rank to the gold term or both models result in a tie.

often achieves better average ranks, it does not consistently outperform BASE on every individual instance. In some techniques, such as *Reductio ad Hitlerum* and *Appeal to Authority*, BASE retains a slight advantage in win rate despite having worse average ranking performance.

A similar pattern appears across different parts of speech. DAPT is most beneficial for semantically rich and domain-relevant words, especially politically and ideologically salient nouns and adjectives such as *šalis* (country), *tauta* (nation), *prezidentas* (president), and *tautinis* (national). For these lemmas, gold-term ranks improve substantially, and DAPT win rates are consistently higher. In contrast, improvements for verbs are more mixed. DAPT helps with context-dependent verbs such as *kovoti* (to fight), but provides more limited or inconsistent gains for frequent and general-purpose verbs such as *būti* (to be), *sakyti* (to say), and *turėti* (to have). Proper nouns such as *Europa* (Europe) also show weaker or even negative effects, suggesting that such terms may already be sufficiently represented in the BASE model.

The lemma–technique analysis further confirms that the impact of DAPT depends on the interaction between the lexical item and its surrounding context. Some lemmas exhibit strong improvements only within particular techniques. For instance, *reikalauti* (demand) in *Waving the Flag*, *sakyti* (say) in *Whataboutism / Red Herring / Straw Man* and *Appeal to Authority*, and *laikas* (time) in *Uncertainty* show substantial gains in ranking (see Figure 6). However, the win-rate analysis reveals that DAPT does not consistently outperform the BASE model across all such cases. This suggests that certain lemma–technique combinations benefit from strong improvements in specific contextual spans, while in other instances the differences are smaller or the BASE model still provides better rankings.

To illustrate this contextual variation, we consider the lemma *sakyti* (to say) using the same technique, *Whataboutism / Red Herring / Straw Man*, but in different contexts.

Na, bet Lietuvos karo strategai [sako], kad viskas bus gerai, mes laimėsime, todėl einu ramiai sau kasti sniego.

EN: Well, Lithuanian military strategists [say] everything will be fine, we will win, so I will calmly go shovel snow.

(DAPT rank = 64, BASE rank = 1)

Context: In this example, *sakyti* is used in a relatively neutral reporting function, introducing a statement without strong rhetorical or evaluative emphasis on the verb itself. This may contribute to the stronger performance of the BASE model.

Sveikas protas reikalauja dialogo su Maskva ir tai suprastų net pirmos klasės mokinius. Bet tik ne mūsų Užsienio reikalų ministerija. Baisu žiūrėti, kaip didžiosios valstybės, tokios kaip JAV, elgiasi kvailiau nei vaikai. Tiesą [sakant], net vaikai taip kvailai nesielgtų.

EN: Common sense requires dialogue with Moscow, and even a first-grader would understand this. But not our Ministry of Foreign

Affairs. It is frightening to see major powers, such as the USA, behaving more foolishly than children. In fact, even children would not behave this foolishly.

(DAPT rank = 1, BASE rank = 61)

Context: In this example, *sakyti* appears in the form *sakant*, functioning as part of the expression *tiesa sakant* (in fact / frankly speaking). Here, the lemma is embedded in a more evaluative and rhetorically marked construction. The surrounding discourse is also more strongly charged, which may contribute to the substantially higher ranking assigned by DAPT.

These examples illustrate that DAPT does not simply improve all occurrences of a given lemma. Rather, its advantage appears to depend on how salient the lemma is within a particular rhetorical context. In some cases, DAPT assigns a much better rank when the lemma is embedded in an evaluative or rhetorically rich construction, while in other contexts the BASE model may still perform better. This example is therefore illustrative and should not be interpreted as a generalizable conclusion on its own.

Overall, the findings suggest that DAPT provides the greatest benefit for ideologically meaningful, domain-relevant, and context-dependent vocabulary, especially when such vocabulary appears in rhetorically rich propaganda spans. Its advantages are more limited for highly frequent, generic, or already well-represented lexical items. Thus, DAPT appears to improve the model's sensitivity to contextual and rhetorical usage rather than uniformly improving the ranking of all lemma occurrences.

C. Propaganda technique detection: sequence tagging and sentence classification

Table III reports the performance of technique span tagging, while Table IV presents the performance of technique recognition when the task is formulated as sentence classification. Overall, considering both the sequence tagging and sentence classification results, DAPT improves propaganda technique recognition for most techniques across both tasks. However, the effect of DAPT is not uniform and depends on the task formulation, as well as on the linguistic and contextual properties of each technique. From a technique-level perspective, the following patterns can be observed:

- For the techniques *Simplification*, *Doubt*, *Uncertainty*, and *Following Behind*, the DAPT model improves performance in both tasks. The MLM analysis also shows that DAPT generally improves lemma ranking for these techniques. The few cases where the BASE model performs better involve general-purpose lemmas such as *turėti* (to have), *galėti* (can), and *būti* (to be), which reflect more neutral contexts rather than propagandistic cues.
- The most dominant technique, *Emotional Expression*, which relies heavily on emotionally charged language, shows mixed results. In sequence tagging, the BASE

TABLE III

TECHNIQUE-TAGGING PERFORMANCE COMPARISON BETWEEN BASE, DAPT, AND GPT-5.2. BEST TEST F1 PER TECHNIQUE IS SHOWN IN **BOLD**.

Technique	Model	Train			Val			Test		
		P	R	F1	P	R	F1	P	R	F1
Emotional expression	Base	47.43	92.29	62.66	50.50	64.67	56.71	39.32	61.82	48.07
	DAPT	92.58	99.56	95.94	38.69	62.02	47.65	39.33	59.22	47.27
	GPT-5.2	—	—	—	—	—	—	43.28	11.28	17.90
Simplification	Base	28.56	57.36	38.13	23.96	41.51	30.38	21.63	44.66	29.14
	DAPT	28.17	64.42	39.20	23.31	46.36	31.02	24.21	49.53	32.53
	GPT-5.2	—	—	—	—	—	—	17.89	16.78	17.32
Doubt	Base	25.79	82.63	39.31	23.25	51.60	32.06	20.38	53.98	29.59
	DAPT	32.00	88.48	47.00	26.92	53.56	35.83	21.96	54.72	31.35
	GPT-5.2	—	—	—	—	—	—	23.16	8.43	12.36
Whataboutism Red Herring Straw man	Base	4.96	98.63	9.44	6.90	31.57	11.33	7.81	13.26	9.83
	DAPT	18.26	65.00	28.51	6.74	25.34	10.65	8.09	23.50	12.04
	GPT-5.2	—	—	—	—	—	—	10.53	3.77	5.56
Appeal to authority	Base	19.64	93.57	32.47	15.88	29.12	20.56	34.58	19.12	24.62
	DAPT	31.67	95.96	47.62	20.69	26.23	23.13	18.98	30.79	23.48
	GPT-5.2	—	—	—	—	—	—	6.25	3.03	4.08
Waving the flag	Base	18.88	94.40	31.47	9.64	35.95	15.21	23.92	46.36	31.56
	DAPT	22.52	96.89	36.55	9.79	37.44	15.52	27.32	47.82	34.78
	GPT-5.2	—	—	—	—	—	—	18.18	7.27	10.39
Uncertainty	Base	4.97	79.25	9.35	1.84	24.77	3.43	3.23	22.49	5.64
	DAPT	24.70	86.06	38.39	2.64	9.39	4.12	6.06	10.12	7.58
	GPT-5.2	—	—	—	—	—	—	3.59	28.95	6.40
Reductio ad Hitlerum	Base	7.48	13.93	9.73	4.08	28.70	7.14	13.26	41.50	20.10
	DAPT	30.20	96.63	46.02	7.73	26.13	11.93	26.51	40.81	32.14
	GPT-5.2	—	—	—	—	—	—	57.14	16.67	25.81
Following behind	Base	20.55	96.35	33.88	4.77	20.82	7.76	19.94	19.86	19.90
	DAPT	15.93	96.20	27.34	5.26	15.53	7.86	16.79	29.31	21.36
	GPT-5.2	—	—	—	—	—	—	1.74	18.18	3.17

model performs slightly better, whereas in sentence classification DAPT slightly outperforms BASE (BASE = 66.08%, DAPT = 66.79%). The MLM analysis also indicates better ranking performance with DAPT. These results can be explained by the high frequency of this technique in the corpus, which allows both models to learn it relatively well. As a result, the performance differences between the models remain minor.

- The *whataboutism/red herring/straw man* technique achieves the lowest performance across both tasks. This technique is difficult to identify, even for human annotators, and is also relatively rare in the dataset. In sequence tagging, DAPT slightly outperforms BASE, although overall performance remains low for both models. In sentence classification, BASE slightly outperforms DAPT (BASE = 49.36%, DAPT = 49.28%), but the difference is negligible. The MLM analysis also shows no clear advantage for either model, although DAPT improves rankings for some emotionally or politically charged lemmas such as *tauta* (nation), *tautinis* (national), *vaikas* (child), and *šalis* (country). This suggests that techniques with strong class imbalance, low frequency, and weak contextual distinctiveness are inherently difficult to model and may require approaches beyond DAPT.

TABLE IV

SENTENCE CLASSIFICATION PERFORMANCE (ACCURACY AND F1). BEST TEST F1 PER TECHNIQUE IS SHOWN IN **BOLD**.

Technique	Model	Train		Val		Test	
		ACC.	F1	ACC.	F1	ACC.	F1
Emotional expression	Base	76.83	71.38	69.09	65.18	70.84	66.08
	DAPT	76.88	74.86	71.41	67.33	73.14	66.79
	GPT-5.2	—	—	—	—	62.22	52.64
Simplification	Base	98.28	97.27	78.51	61.11	78.23	60.25
	DAPT	76.28	68.03	74.14	62.70	75.90	62.55
	GPT-5.2	—	—	—	—	51.98	43.46
Doubt	Base	82.14	70.71	76.39	61.99	76.18	59.90
	DAPT	80.90	69.01	81.05	64.55	81.05	62.71
	GPT-5.2	—	—	—	—	74.53	52.75
Whataboutism Red Herring Straw Man	Base	85.86	50.73	89.37	53.83	87.75	49.36
	DAPT	96.47	84.16	85.92	51.31	85.89	49.28
	GPT-5.2	—	—	—	—	71.18	45.16
Appeal to authority	Base	75.35	50.97	79.65	52.49	77.08	48.85
	DAPT	98.57	92.14	94.17	62.77	94.01	58.07
	GPT-5.2	—	—	—	—	79.40	50.36
Waving the flag	Base	97.91	80.34	96.29	67.62	96.31	68.85
	DAPT	99.02	90.47	96.89	68.41	96.93	68.47
	GPT-5.2	—	—	—	—	97.05	52.22
Uncertainty	Base	96.60	66.97	97.05	53.90	97.33	50.46
	DAPT	95.79	63.34	95.07	51.53	95.72	55.78
	GPT-5.2	—	—	—	—	95.81	51.73
Reductio ad Hitlerum	Base	99.20	85.23	98.12	61.93	98.76	67.43
	DAPT	99.47	90.04	98.27	67.25	98.82	74.03
	GPT-5.2	—	—	—	—	98.86	77.42
Following behind	Base	97.44	52.20	99.16	51.92	98.67	51.89
	DAPT	98.53	68.10	97.92	55.63	97.92	68.16
	GPT-5.2	—	—	—	—	98.07	59.77

- The *Appeal to Authority* technique shows contrasting results across tasks. In sequence tagging, the BASE model significantly outperforms DAPT (BASE = 27.40%, DAPT = 23.48%), whereas in sentence classification DAPT substantially outperforms BASE (DAPT = 58.07%, BASE = 48.85%). The MLM analysis indicates that the BASE model better ranks general-purpose lemmas such as *būti* (to be), *reikalauti* (to demand), and *sakyti* (to say), while DAPT performs better on more contextually specific lemmas. Since this technique often relies on relatively general language referring to authority figures, the benefits of DAPT appear to be less consistent and more dependent on the task formulation.
- For the *Waving the Flag* technique, DAPT improves performance in the sequence tagging task (DAPT = 34.78%, BASE = 31.56%), while BASE performs slightly better in the sentence classification task (BASE = 68.85%, DAPT = 68.47%). In both cases, the differences are small. Despite being relatively rare in the corpus, this technique achieves high classification performance, likely due to its clear and distinctive patriotic context. Key lemmas include *tauta* (nation), *tautinis* (national), *valstybė* (state), and *šalis* (country). The MLM analysis shows that DAPT

tends to assign higher ranks to these lemmas, but the overall improvements remain modest.

- The *Reductio ad Hitlerum* technique shows improved performance with DAPT compared with BASE in both tasks. However, in sentence classification, the GPT model outperforms both BASE and DAPT (77.42%). Overall, classification performance for this technique is high across all models, while sequence tagging performance remains relatively low. This can be explained by the technique's highly distinctive context, which benefits sentence-level classification. In contrast, sequence tagging requires precise span boundary detection, which is more difficult, especially given the low frequency of this technique in the corpus.

These task- and technique-level differences suggest that the benefits of DAPT are not uniform across techniques and depend on the modeling approach. In the sequence tagging task, performance is strongly influenced by the tagging scheme and by the model's ability to identify precise span boundaries. In our case, since binary sequence tagging is used, the model is evaluated based on whether it correctly identifies propaganda spans, regardless of the specific technique label. The results show that DAPT mainly improves recall, while precision remains relatively low. This suggests that the model is often able to recover characters or tokens that belong to gold propaganda spans, but it also tends to include tokens outside the annotated boundaries.

Therefore, the relatively limited improvement in precision indicates that, in the sequence tagging setting, DAPT primarily helps the model recover a larger proportion of true propaganda instances rather than improving its ability to predict exact span boundaries. In contrast, in the sentence classification task, the model is not required to identify span boundaries. As a result, the classification results more directly reflect the model's ability to recognize propaganda techniques at the sentence level.

The classification results also suggest that class frequency is not the only factor determining performance. Techniques with distinctive and separable contexts, such as *Reductio ad Hitlerum* and *Waving the Flag*, can be recognized relatively well even when they have fewer training instances. At the same time, rare techniques with less distinctive or more ambiguous contexts remain difficult to classify.

Compared with GPT, DAPT achieves better performance across almost all techniques and tasks, with the only exception being *Reductio ad Hitlerum* in sentence classification (GPT = 77.42%, DAPT = 74.03%). This exception is likely due to the highly distinctive context of this technique, which makes it easier to recognize even without task-specific supervised training. The results also show clear differences between supervised sequence tagging models and GPT. BASE and DAPT generally achieve higher recall than precision, suggesting that they identify more gold-standard tokens but often overtag surrounding tokens. In contrast, GPT achieves higher precision and lower recall for most techniques, indicating that it produces fewer predictions but focuses more on complete and coherent span boundaries. This difference reflects the modeling formulation: supervised sequence tagging labels individual tokens, whereas

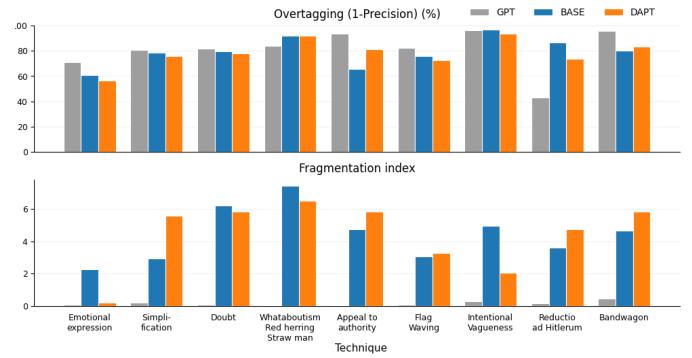


Fig. 8. Overtagging rate and fragmentation index for GPT, DAPT, and BASE in the sequence tagging task.

GPT autoregressively generates textual span annotations based on the input and prompt. Therefore, GPT can produce more coherent spans, but it tends to miss more valid propaganda instances.

Overall, the main conclusion is that DAPT improves propaganda technique recognition in many cases, but its effectiveness depends on both the task formulation and the contextual distinctiveness of the technique. One possible explanation for cases where DAPT does not improve performance, or slightly reduces it, is that domain adaptation may shift the model toward lexical and contextual patterns that are frequent in the non-credible news corpus but not necessarily discriminative for a specific propaganda technique. This may be problematic for techniques expressed through general-purpose language, such as *Appeal to Authority*, or for techniques with weak contextual distinctiveness, such as *Whataboutism / Red Herring / Straw Man*. In such cases, DAPT may increase sensitivity to domain-specific discourse while reducing the model's reliance on broader semantic cues learned during general pretraining.

D. Analysis of Annotation Quality: GPT, BASE, and DAPT

To further analyze the qualitative behavior of the models in the sequence tagging formulation, we measured two additional annotation-level metrics:

Overtagging rate represents the proportion of predicted characters that do not overlap with any gold-standard annotation. This metric captures the extent to which a model overextends span boundaries or assigns a technique label to irrelevant text. In general, this corresponds to 1 - precision.

Fragmentation index quantifies how many predicted spans overlap a single gold span on average. Values close to one indicate coherent one-to-one span alignment, whereas higher values suggest that the model splits gold spans into multiple smaller fragments.

Figure 8 presents the results for the sequence tagging task on the test data, showing the overtagging rate and fragmentation index for the GPT, DAPT, and BASE models. From the perspective of overtagging, the differences between the models are small. All models exhibit relatively high overtagging rates, typically ranging between 70% and 90% across techniques, indicating that they tend to label too many characters beyond the gold-standard span boundaries. Since this pattern is observed

	GOLD-ANNOTATION	GPT ANNOTATION	BASE ANNOTATION	DAPT ANNOTATION
All Techniques	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Emotional Expression	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Doubt	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Simplification	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Intentional vagueness	Not in gold annotation	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Appeal to Authority	Not in gold annotation	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Bandwagon	Not in gold annotation	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naujausias mano įrašas pagal šią savaitę aktualiją. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų, iš tų kurie nevisada sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.

Fig. 9. Annotations produced by BASE, DAPT, and GPT for the case study article.

across all models and is not substantially affected by DAPT, overtagging appears to be a general sequence tagging issue rather than a model-specific problem. One possible explanation is the use of a binary tagging scheme, which does not explicitly focus on span boundary information.

Regarding the fragmentation index, GPT consistently obtains much lower values than BASE and DAPT. This indicates that GPT tends to annotate more continuous and coherent spans, whereas BASE and DAPT frequently split a single gold span into multiple smaller predicted spans. This pattern is plausible given that BASE and DAPT are token-level models. Since tokenization may split words into subword units, partial-word labeling can occur, for example when only part of a word is assigned a technique label. This leads to fragmented annotations and, consequently, higher fragmentation scores. These results further confirm that the influence of DAPT in tasks such as sequence tagging may depend on the selected tagging scheme or, more generally, on the task itself.

Case Study: Model Annotations on a Test Article

Additionally, we conducted a qualitative analysis of a single test article annotated by the GPT, BASE, and DAPT models. The article text is provided below.

Naujausias mano įrašas pagal šią savaitę aktualijas. Vaikai tampa preke kuri jau dabar yra nusavinama iš mažiau turtingų šeimų. Iš tų kurie nevisada sugėba pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000–70 000 šeimų kurios negali susilaukti vaikų. Vaikų globėjų institucija virsta

tam tikra verslo šaka kuri suinteresuota, jog darbo nepritrūktų. Šalia viso to yra homoseksualių partnerystė kuri natūraliai negali susilaukti vaikų, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.

EN: My latest post is about this week's current issue. Children are becoming a commodity that is already being taken away from less wealthy families— from those who are not always able to resist. It should be kept in mind that in Lithuania alone there are around 50 000–70 000 families who are unable to have children. The child guardianship institution is turning into a kind of business sector that is interested in ensuring there is no shortage of work. In addition to all this, there is homosexual partnership, which naturally cannot have children, as well as the possibility of pedophilia among potential guardians.

Figures 9 and 10 illustrate the annotations for this example article. Figure 9 shows the sequence tagging results, while Figure 10 presents the sentence classification results.

In the gold annotation, the article contains three techniques: *Emotional Expression*, *Doubt*, and *Simplification*. The *Emotional Expression* label covers almost the entire article, as the text discusses sensitive topics such as child protection, homosexuality, and pedophilia using emotionally charged language. The *Doubt* label is associated with statements that undermine trust in child protection institutions, while *Simplification* is assigned to the sentence that reduces a complex institutional issue to a profit-driven explanation.

	GOLD ANNOTATION	GPT ANNOTATION	BASE ANNOTATION	DAPT ANNOTATION
All Techniques	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Emotional Expression	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Doubt	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Simplification	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.
Appeal to Authority	Not in gold annotation	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.	Naudojamas mano įrašas pagal šią sąveikų aktualiją. Vakiai tampa preke kuri jau dabar yra nusavinama iš mažiaus turtingų šeimų, iš tų kurie nevaido sugėta pasipriešinti. Reikia turėti omenyje, jog vien Lietuvoje yra apie 50 000 - 70 000 šeimų kurios negali susilaikyti vaikių. Vaiški globos įstaigose virsta tam tikra versio šaka kuri suinteresuota, jog darbo nepriimtų. Sala viso to yra homoseksuali partnerystė kuri natūraliai negali susilaikyti vaikių, o taip pat ir pedofilijos tikimybė tarp potencialių globėjų.

Fig. 10. Sentence-level annotations produced by BASE, DAPT, and GPT for the case study article.

The GPT annotations show that the output depends strongly on the task formulation. In the sequence tagging setting, GPT assigned *Emotional Expression*, *Doubt*, and *Simplification* to the final sentence. However, in the sentence classification setting, only *Doubt* was assigned to this sentence. Compared with the gold annotation, GPT therefore over-assigned *Simplification* in the sequence tagging task and missed *Emotional Expression* in the sentence classification task. In both settings, GPT also struggled to identify *Simplification* correctly.

The BASE models identified *Emotional Expression* better than GPT and DAPT in this example. Both the sequence tagging and sentence classification models correctly detected the same two sentences for this technique. However, BASE also showed a tendency to over-assign labels. In particular, both BASE models assigned *Simplification* to a large part of the text, and they also predicted techniques that were not present in the gold annotation, such as *Appeal to Authority*. The sequence tagging model additionally assigned *Bandwagon* and *Intentional Vagueness* incorrectly.

The DAPT models behaved more conservatively. Unlike GPT and BASE, they did not assign techniques that were completely absent from the gold annotation. However, this also led to missed annotations. For example, the DAPT sentence classification model did not identify *Doubt* at all, while the sequence tagging model detected only a few isolated words belonging to this technique. DAPT also under-assigned *Simplification*, suggesting that this technique remained difficult to distinguish from *Emotional Expression*.

This example provides an illustrative case of the qualitative differences between the models. GPT produces more coherent spans in the sequence tagging setting, although its annotations vary depending on the prompting format. BASE captures *Emotional Expression* relatively well but tends to over-assign techniques. DAPT is more conservative and avoids assigning completely unrelated techniques, but it may miss

some valid labels. The example also confirms that sequence tagging models often produce fragmented annotations, which is consistent with the fragmentation index analysis.

VII. CONCLUSION

In this study, we investigated the impact of domain-adaptive pretraining (DAPT) on fine-grained propaganda technique detection in Lithuanian, a low-resource language. Using LT-MLKM-modernBERT as the backbone model, we continued pretraining on 63,648 domain-specific Lithuanian news articles from non-credible outlets. We then evaluated the effect of DAPT through masked language modeling analysis, supervised propaganda technique recognition under sequence tagging and sentence classification formulations, comparison with a zero-shot GPT-5.2 baseline, and a qualitative analysis of annotation behavior.

The results can be summarized as follows:

- Overall, DAPT improves propaganda technique detection in Lithuanian. For both sequence tagging and sentence classification tasks, DAPT improves results for 7 out of 9 techniques; however, the improvements occur for different techniques, indicating that its impact varies across techniques and task formulations.
- In sequence tagging, DAPT improves performance mainly by increasing recall, while precision remains relatively low. This suggests that DAPT helps the model recover more gold-standard propagandistic spans, but exact span boundary detection remains challenging.
- Sentence classification provides a clearer view of DAPT's impact, as it is not affected by span boundary detection. The results show more consistent improvements and indicate that DAPT significantly improves propaganda technique classification. In the two techniques where the BASE model performed better (Whataboutism/Red Herring/Straw Man and Waving the Flag), the difference

between the base and DAPT models was very small, less than 0.4%. For the remaining techniques, DAPT improved performance by between 2.3% and 16.27%.

- The classification results also suggest that class balance is not the only factor influencing technique identification: techniques with distinctive and separable contexts can be recognized relatively well even with fewer training instances, while the most difficult techniques are those that are both rare in the corpus and have less distinctive contexts.
- The MLM analysis shows that DAPT mainly improves the model's ability to rank propaganda vocabulary that depends on context more highly. The strongest improvements are observed for politically and ideologically meaningful nouns and adjectives. However, DAPT provides weaker gains for highly frequent and general vocabulary, suggesting that it primarily improves sensitivity to contextual usage within the propaganda domain rather than overall lexical prediction performance.
- The annotation quality analysis shows that DAPT produces more conservative annotations than BASE at both the sentence and span levels. However, for span-level annotation, DAPT does not substantially resolve span-boundary errors, and overtagging remains a common issue for both BASE and DAPT. The fragmentation analysis further shows that both models frequently produce fragmented spans. This confirms that the binary sequence tagging scheme is problematic, as models classify individual tokens without explicit boundary information, which can result in isolated tokens being tagged instead of complete propaganda spans.

Taken together, these findings suggest that domain-adaptive pretraining is particularly effective when propaganda techniques rely on domain-specific context. Rather than uniformly improving performance, DAPT enhances the model's sensitivity to contextual and semantic cues that are characteristic of propagandistic discourse. This highlights the importance of domain adaptation in low-resource languages, where general-domain pretraining may not provide enough exposure to specialized propaganda-related discourse.

This study contributes to domain adaptation research by showing how DAPT affects both lexical prediction and downstream propaganda technique classification in a low-resource language. It also contributes to propaganda detection research by demonstrating that domain adaptation improves the recognition of context-dependent propaganda techniques. From a practical perspective, the findings suggest that DAPT-based supervised models can support propaganda monitoring in low-resource languages, although span-boundary errors and fragmented annotations indicate that such systems should be used as decision-support tools rather than fully automatic annotation systems.

A. Limitations

One of the main limitations of this work is the use of a binary token classification approach for technique tagging. This is reflected in the fragmentation index results, which show that

the fine-tuned supervised models produce significantly higher fragmentation scores compared to the GPT-5.2 model. These results indicate that supervised models tend to split a single gold-standard span into multiple smaller spans. This behavior is also evident in the qualitative annotation examples, where fine-tuned models sometimes assign techniques to individual words rather than complete spans. This can be attributed to the nature of binary token tagging and subword tokenization, where words may be divided into smaller units, leading in some cases to partial-word labeling.

Another limitation concerns the imbalance of the HALT-PROP dataset. Several techniques are sparsely represented, resulting in limited training data for reliable learning. However, the imbalance reflects natural distribution patterns rather than sampling bias. This highlights the need for more robust training strategies to better support the detection of rare propaganda techniques.

ACKNOWLEDGMENT

The first author would like to thank NordForsk and the Research Council of Lithuania (LMTLT), agreement No [SNORDFORSK-26-5] for support through the funding to "Enhancing resilience against biological threats in the Nordic and Baltic countries (BIO-NORD)", project number 233158. The second author would like to thank, in addition to NordForsk BIO-NORD project, Helsingin Sanomat foundation for support through the funding to "Addressing the Technological and Societal Challenges of Disinformation: Multimodal Verification Tools for Advancing the Impact of Fact-Checking". We also thank Dr. Giulio Corsi from the University of Cambridge for reviewing the manuscript and providing valuable comments.

REFERENCES

- [1] J. Jose, C. Geeng, K. O. Morales, D. McCoy, and R. Greenstadt, "What's in a Label? Propaganda Labels and User Sharing Behavior on Social Media Platforms," In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 19, no. 1, pp. 918–934, 2025, doi: 10.1609/icwsm.v19i1.35853.
- [2] G. Da San Martino, A. Barrón-Cedeño, and P. Nakov, "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*, pp. 162–170, Hong Kong, China, Association for Computational Linguistics, 2019, doi: 10.18653/v1/D19-5024.
- [3] G. Da San Martino, A. Barrón-Cedeño, H. Wachsmuth, R. Petrov, and P. Nakov, "SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval-2020)*, Barcelona (online), International Committee for Computational Linguistics, 2020, pp. 1377–1414, doi: 10.18653/v1/2020.semeval-1.186.
- [4] D. Dimitrov, B. Bin Ali, S. Shaar, F. Alam, F. Silvestri, H. Firooz, P. Nakov, and G. Da San Martino, "SemEval-2021 Task 6: Detection of Persuasion Techniques in Texts and Images," In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pp. 70–98, Online, Association for Computational Linguistics, 2021, doi: 10.18653/v1/2021.semeval-1.7.
- [5] J. Piskorski, N. Stefanovitch, G. Da San Martino, and P. Nakov, "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)*, pp. 2343–2361, Toronto, Canada, Association for Computational Linguistics, 2023, doi: 10.18653/v1/2023.semeval-1.317.

- [6] D. Dimitrov, F. Alam, M. Hasanain, A. Hasnat, F. Silvestri, P. Nakov, and G. Da San Martino, "SemEval-2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes," In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pp. 2009–2026, Mexico City, Mexico, Association for Computational Linguistics, 2024, doi: 10.18653/v1/2024.semeval-1.275
- [7] F. Alam, H. Mubarak, W. Zaghouani, G. Da San Martino, and P. Nakov, "Overview of the WANLP 2022 Shared Task on Propaganda Detection in Arabic," In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pp. 108–118, Abu Dhabi, United Arab Emirates (Hybrid), Association for Computational Linguistics, 2022, doi: 10.18653/v1/2022.wanlp-1.11
- [8] J. Piskorski, D. Dimitrov, F. Dobranić, M. Ernst, J. Haneczok, I. Koychev, N. Ljubešić, M. Marcinczuk, A. Modzelewski, I. Moravski, and R. Yangarber, "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 (Slavic NLP 2025)*, pp. 254–275, Vienna, Austria, Association for Computational Linguistics, 2025, doi: 10.18653/v1/2025.bsnlp-1.27
- [9] S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N. A. Smith, "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8342–8360, Online, Association for Computational Linguistics, 2020, doi: 10.18653/v1/2020.acl-main.740
- [10] T. Caselli, V. Basile, J. Mitrović, and M. Granitzer, "HateBERT: Retraining BERT for Abusive Language Detection in English," In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pp. 17–25, 2021, doi: 10.18653/v1/2021.woah-1.3
- [11] A. Paraschiv, D. C. Cercel, and M. Dascalu, "UPB at SemEval-2020 Task 11: Propaganda Detection with Domain-Specific Trained BERT," In *Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval-2020)*, pp. 1853–1857, Barcelona, Spain (Online), Association for Computational Linguistics, 2020, doi: 10.18653/v1/2020.semeval-1.244
- [12] T. Chavan and A. Kane, "Large Language Models for Multi-Label Propaganda Detection," arXiv preprint arXiv:2210.08209, 2022, doi: 10.48550/arXiv.2210.08209
- [13] J. Piskorski, A. Jorge, M. D. P. Silvano, N. Guimarães, A. F. Pacheco, and N. Yu, "Overview of the CLEF-2024 CheckThat! Lab Task 3 on Persuasion Techniques," In *Working Notes of CLEF 2024 – Conference and Labs of the Evaluation Forum*, 2024, doi: 10.1007/978-3-031-71908-0_2
- [14] Y. Liu, X. F. Zhang, D. Wegsman, N. Beauchamp, and L. Wang, "POLITICS: Pretraining with Same-Story Article Comparison for Ideology Prediction and Stance Detection," In *Findings of the Association for Computational Linguistics: NAACL 2022*, pp. 1354–1374, 2022, doi: 10.18653/v1/2022.findings-naacl.101
- [15] T. H. Fahim and K. A. Khan, "PerceptionLab at BLP-2025 Task 1: Domain-Adapted BERT for Bangla Hate Speech Detection: Contrasting Single-Shot and Hierarchical Multiclass Classification," In *Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025)*, pp. 498–507, 2025, doi: 10.18653/v1/2025.banglalp-1.45
- [16] I. Rizgeliėnė, V. Zubaitienė, N. Maliukevičius, et al., "HALT-PROP: Human-Annotated Lithuanian Textual Corpus for Propaganda Narratives and Techniques," *Scientific Data*, vol. 13, p. 47, 2026, doi: 10.1038/s41597-025-06367-w.
- [17] M. Hasanain, M. A. Hasan, F. Ahmad, R. Suwaileh, M. R. Biswas, W. Zaghouani, and F. Alam, "ArAIEval Shared Task: Propagandistic Techniques Detection in Unimodal and Multimodal Arabic Content," In *Proceedings of the Second Arabic Natural Language Processing Conference*, pp. 456–466, Bangkok, Thailand, Association for Computational Linguistics, 2024, doi: 10.18653/v1/2024.arabicnlp-1.44
- [18] P. Moral, G. Marco, J. Gonzalo, J. Carrillo-de-Albornoz, and I. Gonzalo-Verdugo, "Overview of DIPROMATS 2023: Automatic Detection and Characterization of Propaganda Techniques in Messages from Diplomats and Authorities of World Powers," *Procesamiento del Lenguaje Natural*, vol. 71, pp. 397–407, 2023, doi: 10.26342/2023-71-31
- [19] M. Z. Ali, B. Pfahringer, and T. Smith, "Fake News Classification in Urdu: A Domain Adaptation Approach for a Low-Resource Language," arXiv preprint arXiv:2512.22778, 2025, doi: 10.48550/arXiv.2512.22778
- [20] Ibrahim Abu Farha, Wajdi Zaghouani, and Walid Magdy. 2021. Overview of the WANLP 2021 Shared Task on Sarcasm and Sentiment Detection in Arabic. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 296–305, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.
- [21] Muhammad Umar Salman, Asif Hanif, Shady Shehata, and Preslav Nakov. 2023. Detecting Propaganda Techniques in Code-Switched Social Media Text. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 16794–16812, Singapore. Association for Computational Linguistics, doi: 10.18653/v1/2023.emnlp-main.1044
- [22] P. N. Ahmad, A. M. Shah, J. Guo, and Y. Liu, "ProST: Spotting propaganda span and technique classification in news articles," *Aslib Journal of Information Management*, vol. ahead-of-print, no. ahead-of-print, 2025, doi: 10.1108/AJIM-08-2024-0660.
- [23] A. M. Ahmad, A. M. Shah, J. Lee, and A. Muhammad, "Misinformation detection on online social networks using pretrained language models," *Information Processing & Management*, vol. 63, no. 1, 2025, Art. no. 104342. doi: https://doi.org/10.1016/j.ipm.2025.104342
- [24] J. D. Krieger, T. Spinde, T. Ruas, J. Kulshrestha, and B. Gipp, "A Domain-Adaptive Pre-Training Approach for Language Bias Detection in News," In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*, pp. 1–7, 2022, doi:10.1145/3529372.3530932
- [25] P. Moral, J. M. Fraile, G. Marco, A. Peñas, and J. Gonzalo, "Overview of DIPROMATS 2024: Detection, Characterization and Tracking of Propaganda in Messages from Diplomats and Authorities of World Powers," *Procesamiento del Lenguaje Natural*, vol. 73, pp. 347–358, 2024.
- [26] State Digital Solutions Agency (SDSA), "LT-MLKM-modernBERT: Lithuanian ModernBERT Language Model," *Hugging Face*, 2025. [Online]. Available: https://huggingface.co/VSSA-SDSA/LT-MLKM-modernBERT. Developed by Vytautas Magnus University (VMU), UAB Neurotechnology, UAB Tilde informacinės technologijos, and MB Krilas.
- [27] P. N. Ahmad, J. Guo, N. M. AboElenein, et al., "Hierarchical graph-based integration network for propaganda detection in textual news articles on social media," *Scientific Reports*, vol. 15, Art. no. 1827, 2025. doi: https://doi.org/10.1038/s41598-024-74126-9
- [28] P. N. Ahmad, L. Yuanchao, K. Aurangzeb, et al., "Semantic web-based propaganda text detection from social media using meta-learning," *Service Oriented Computing and Applications*, 2024. doi: https://doi.org/10.1007/s11761-024-00422-x

Ieva Rizgeliėnė is a PhD candidate and junior researcher in computer science at the Institute of Data Science and Digital Technologies, Faculty of Mathematics and Informatics, Vilnius University, Lithuania. Her primary research interests lie at the intersection of artificial intelligence, natural language processing (NLP), and computational linguistics, with a particular focus on detecting propaganda and disinformation.



Hannele Seeck is Professor of Communication Sciences at LUT University, Finland, where she leads the Disinformation, Propaganda & Soft Power Research Lab (DPS-Lab). Her Lab's research examines ideological influence, disinformation, and global risks, with a particular focus on East Asia and the Global South. She is a graduate (MSc, PhD and Post-doc) of the London School of Economics and Political Science (LSE), where she is also currently a Visiting Professor in the Department of Media and Communications. At the University of Cambridge, she is an Affiliated Researcher at the Centre for the Study of Existential Risk (CSER), a Research Fellow (Honorary Appointment) at Judge Business School, and a Life Member of Clare Hall. She has published some 120 academic and professional works, including books, and articles in journals such as *Journalism Studies*, *Media, Culture & Society*, and *Organization*.

