Data in Brief 67 (2026) 113000

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A dataset for detecting emotional manipulation techniques in Lithuanian text

Rita Butkienė*, Algirdas Šukys, Edgaras Dambrauskas, Voldemaras Žitkus, Linas Ablonskis, Evaldas Vaičiukynas, Paulius Danėnas, Rimantas Butleris

Centre of Information Systems Design Technologies, Faculty of Informatics, Kaunas University of Technology, K. Baršausko st. 59, 51368423, Kaunas, Lithuania

ARTICLE INFO

Article history:
Received 25 March 2026
Revised 19 June 2026
Accepted 19 June 2026
Available online 25 June 2026

Dataset link: Lithuanian Emotional Manipulation Dataset (Original data)

Keywords:
Manipulative technique detection
Manipulative span identification
Emotional manipulation
Dataset annotation
Prompt engineering
GPT

ABSTRACT

This paper presents a dataset of Lithuanian language comments annotated with emotional manipulation techniques. The source material comprises comments from Lithuanian news portal texts. A total of 1000 comments were selected and manually annotated by four human annotators using Label Studio. The annotators identified text fragments corresponding to fourteen emotional manipulation techniques, producing span-level human annotations that form the primary component of the dataset.

In addition to the manual annotations, the dataset includes machine-generated outputs created by the GPT-4.1 model. These outputs were produced by executing prompts designed for the detection and classification of emotional manipulation span. Several prompting strategies were applied, and each prompt was run five times to capture variability in model behaviour. Because GPT-4.1 does not always return verbatim text excerpts, all generated spans were post-processed to extract precise fragments from the original comments and compute their character-level offsets.

The resulting dataset encompasses four components: unprocessed source comments, human-annotated spans, prompt templates, and GPT-4.1 generated predictions.

The dataset provides the first publicly available Lithuanian resource annotated for emotional manipulation techniques

* Corresponding author.
E-mail address: rita.butkiene@ktu.lt (R. Butkienė).

https://doi.org/10.1016/j.dib.2026.113000

and can support research on manipulation and persuasion phenomena in a morphologically rich, low-resource language. The human annotations may be used as a benchmark for evaluating computational models for span extraction and technique classification. The inclusion of prompts and corresponding GPT-4.1 outputs enables the study of large language model behaviour under different prompting strategies and facilitates prompt engineering research without repeating inference runs. The dataset may also be reused for training or evaluating multilingual and cross-lingual models and offers a foundational framework for the development of annotation schemes and guidelines in related corpus construction efforts.

Specifications Table

Subject	Computer Sciences
Specific subject area	Detection of emotional manipulation techniques in text, Natural language processing, Machine learning, Text annotation using large language models, Prompt engineering.
Type of data	Table: CSV, JSON. Raw, Generated, Processed.
Data collection	The source material originates from comments from Lithuanian news portals in the LITIS corpus [1] and the *Facebook* pages of news portals. A total of 1000 comments were selected and manually labelled (using Label Studio [2]) by four human annotators, marking fragments that represent emotional manipulation techniques. Generated data are GPT-4.1 outputs: spans produced when executing manipulative span identification and classification prompts, replicated over five iterations. Because GPT-4.1 does not always return exact excerpts, all generated spans were post-processed by extracting precise fragments from the original text and computing their positions.
Data source location	Primary source of comments is Lithuanian news portals *Delfi.lt* and *Lrytas.lt*, and the *Facebook* pages of news portals. The secondary source of raw data is the LITIS corpus [1], stored in the CLARIN-LT repository,[1] licensed under Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0). Location of selected and annotated comments, and generated data is Kaunas University of Technology, Kaunas, Lithuania.
Data accessibility	Repository name: **Lithuanian Emotional Manipulation Dataset** Data identification number: https://doi.org/10.5281/zenodo.20053557 Direct URL to data: https://doi.org/10.5281/zenodo.20053557
Related research article	R. Butkienė, A. Šukys, E. Dambrauskas, L. Ablonskis, E. Vaičiukynas, T. Danėnas, R. Butleris. Using LLMs for Pre-Annotation of Emotional Manipulation Techniques in a Low-Resource Language Corpus: Are We There Yet?. Applied Sciences, In Press.

[1] https://clarin.vdu.lt/xmlui/?locale-attribute=lt.

1. Value of the Data

- The dataset provides Lithuanian-language comments posted by readers of news articles and annotated at the span level for 14 emotional manipulation techniques. It enables fine-grained analysis of manipulative and persuasive language in user-generated content, a textual domain that differs substantially from edited news articles and is unrepresented in existing Lithuanian resources.
- The dataset combines human-annotated gold-standard spans with automatically generated annotations produced by the GPT-4.1 model, supporting comparative evaluation of manual and large language model–based emotional manipulation detection in a low-resource lan-

guage.The dataset includes 30 distinct prompt variant templates developed using four prompt engineering strategies, together with results from five repeated executions at temperature settings 0 and 1, allowing analysis of model behaviour, output variability, and instruction sensitivity without re–executing computationally expensive prompt runs.

- The dataset is accompanied by Python scripts that enable full reproducibility of the LLM annotation workflow and facilitate experimentation with new prompt formulations and configurations.
- The dataset enables a diverse set of downstream tasks and benchmarking scenarios, including:
    - Span extraction: evaluating models on detecting exact spans of emotional manipulation;
    - Technique classification: measuring classification performance independently of span detection;
    - End-to-end detection: evaluating complete pipelines – joint span detection and technique classification;
    - Human–LLM agreement analysis: quantifying alignment between human annotations and LLM outputs;
    - Prompt engineering evaluation: evaluating how prompting strategies affect LLM performance;
    - Augmentation material: expanding future datasets, especially if they contain some overlap with emotional manipulation techniques annotated here;
    - Cross-lingual transfer: testing generalization from high-resource languages.
- The dataset could be used for practical deployment applications, including:
    - Content moderation systems: detecting manipulative or emotionally charged language in online discussions;
    - Disinformation and influence monitoring: identifying persuasive tactics in public discourse;
    - Educational tools and media literacy platforms: helping users recognize emotional manipulation techniques;
    - Decision-support systems for journalists and analysts: highlighting potentially manipulative reader responses;
    - Human-in-the-loop moderation pipelines: where automatically detected spans assist expert review.
- Although designed for Lithuanian, the dataset is relevant for multilingual and cross-lingual emotional manipulation research. Emotional manipulation strategies are conceptually transferable across languages, while Lithuanian's morphological richness and low-resource characteristics make the dataset a valuable testbed for evaluating model generalization.
- A limitation of the gold standard dataset is the imbalance in the distribution of emotional manipulation techniques, with some techniques occurring substantially more frequently than others. This imbalance should be considered when using the data for model training, validation, or prompt evaluation.

2. Background

This dataset was compiled to support research on the detection of emotional manipulation techniques in Lithuanian language comments published on news portals or social media. The motivation for creating this resource stems from the need to research computational methods for detecting manipulative language strategies commonly used for persuasive or propaganda purposes. Such methods rely on the availability of systematically annotated data; however, manual creation of large, high-quality corpora is costly and time-consuming. These challenges are particularly acute for Lithuanian, a low-resource language with limited natural language processing assets.

To address this challenge, large language models were explored as a viable means for augmenting the annotation workflows [3,4]. Because the output of such models strongly depends on both model settings and prompt formulation [5], experimenting with different prompt designs is necessary to capture variability in model behaviour and performance. Conducting such experiments requires repeated executions and can be computationally expensive.

The resulting dataset comprises 1000 comments annotated both manually and automatically. The newly introduced components are: (1) manual annotations forming a gold-standard reference, and (2) automatic annotations generated using the GPT-4.1 model through 30 prompts designed according to four distinct prompting strategies and executed repeatedly under different temperature settings.

To the best of our knowledge, the only other publicly available Lithuanian dataset annotating manipulation techniques, HALT-PRO [6], was developed in parallel to this work. HALT-PRO comprises 1000 texts, annotated with 10 propaganda techniques, and a total of 13,591 annotated spans. It employs a different taxonomy of techniques than our dataset, although there is partial overlap: both datasets include techniques such as Flag-Waving, Bandwagon, and Reduction ad Hitlerum/Stalinum. Other HALT-PRO categories, such as Doubt/Smears and Emotional Expression, correspond in our scheme to a broader set of more fine-grained emotional manipulation techniques. Unlike HALT-PRO, which focuses on propaganda narratives in news articles, our dataset uniquely provides span-level annotations of emotional manipulation techniques in Lithuanian user-generated comments, addressing a complementary and underexplored textual domain.

3. Data Description

The dataset is available in the Zenodo dataset repository [7], the version matching this paper is v3. Previous versions track changes made during the review process of this paper. A conceptual structure of the dataset is presented in Fig. 1. It contains: a) comments used for annotation; b) the gold dataset, which consists of 14 files including manually identified fragments (spans) of emotional manipulation instances in the comments; c) the generated dataset with spans of emotional manipulation instances, identified in comments using GPT-4.1 and 4 different prompt engineering strategies and temperature settings.

The physical structure of the dataset is presented in Fig. 2. There are three top-level directories: *data, prompts,* and *scripts*. The *data* directory stores the above-mentioned data: a) the comments are stored in the file *data/comments.csv*; b) the gold dataset is stored in the *CSV* files of the subdirectory *data/gold_dataset,* where files are named by the technique, e.g., *doubt.csv*, and c) the prompt-generated dataset is stored in the *CSV* files of subdirectories dedicated to 4 prompting strategies (*data/strategy_I, data/strategy_II*, etc.). Files in the prompt-generated dataset are named *spans.csv*. The *prompts* directory contains all prompt templates used in each prompt engineering strategy. It is organized into subdirectories corresponding to each prompting strategy (e.g., *prompts/strategy_I/, prompts/strategy_II/*, etc.). Each strategy-specific subdirectory includes prompt template files in both Lithuanian and English, enabling multilingual experimentation and reproducibility (e.g., *prompt_lt.json and prompt_en.json).* The *scripts* directory contains Python scripts for performing emotional language identification with the OpenAI API and post-processing procedures such as span alignment. This directory also contains the *output* directory, which is empty in the dataset and is used solely to store the results of emotional language identification with GPT when reproducing our experiment.

3.1. Comments

A total of 1000 comments are archived in the file *data/comments.csv*. A major part of the comments was selected from the LITIS corpus [1] whose license allows us to share (copy and redistribute the material in any medium or format) and to adapt (remix, transform, and build upon the material) its content This corpus includes comments sourced from two Lithuanian

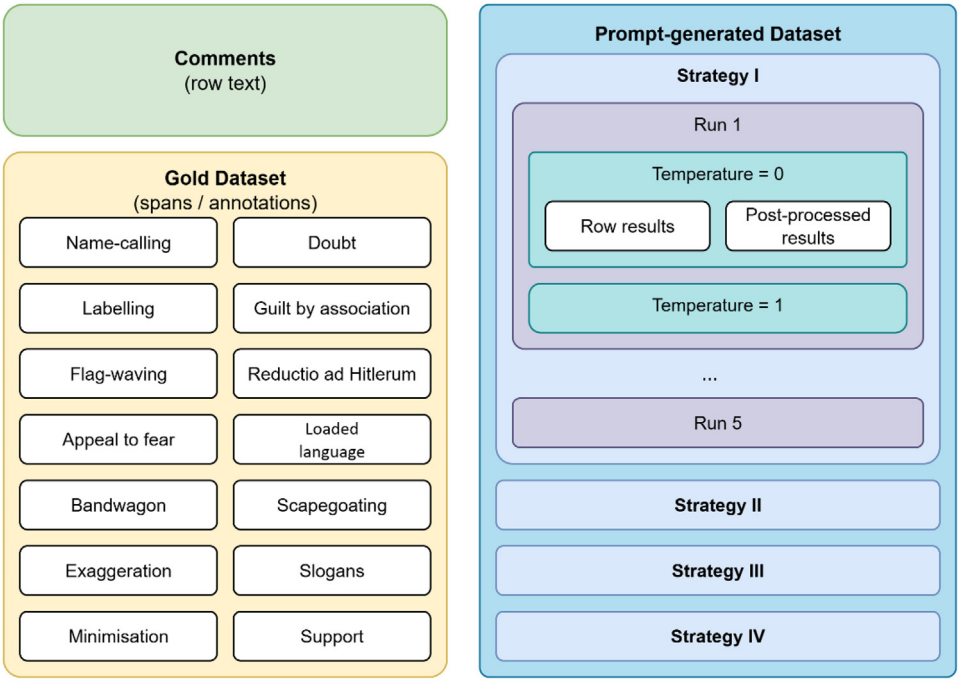

Fig. 1. The conceptual structure of the dataset.

Table 1
Information on dataset sources.

Source	#Comments
Delfi.lt (Sourced from LITIS corpus)	892
Lrytas.lt (Sourced from LITIS corpus)	3
Facebook pages of news portals	105

news portals, selecting news articles published between 2010 and 2014. The comments are written by anonymous visitors. Comments from the LITIS corpus were selected in two stages: (1) a subset of comments intended for the construction of a hate speech dataset was selected using keyword-based searches targeting potential objects of hate speech (e.g., race, nationality, gender, sexual orientation, political views, etc.); (2) this subset was manually reviewed to select those representing manipulative content. The emotional manipulation corpus was further expanded with highly engaging *Facebook* discussions posted between 2021 and 2025 on the pages of major news outlets, focusing on articles that received at least 100 comments.

All comments included in the dataset are publicly available, and their collection complies with LITIS corpus and other source terms of use. To address ethical and legal considerations, the data were fully anonymized, and no personally identifiable information was retained. The dataset is intended solely for research and educational purposes, focusing on the analysis of emotional manipulation techniques rather than the identification or profiling of individual users. The structure of the dataset by sources is presented in Table 1. For comments originating from the LITIS corpus, the original source can be identified by searching for the comment text within the corpus.

The structure of the comments file and the examples is presented in Table 2.

```
|- data/
|   |- strategy_I
|              |- run_1
|                  |- temp_0
|                     |- post-processed
|                        |- spans.csv
|                     |- raw
|                        |- spans.csv
|                  |- temp_1
|                     |- post-processed
|                        |- spans.csv
|                     |- raw
|                        |- spans.csv
|            ...
|              |- run_5
|                  |- ... (same base structure as run_1)
|   ...
|   |- strategy_IV
|              |- ... (same base structure as strategy_I)
|   |- gold_dataset
|              |- appeal_to_fear.csv
|              |- ... (files for the remaining techniques)
|   |- comments.csv
|- prompts/
|   |- strategy_I
|              |- prompt_lt.json
|              |- prompt_en.json
|   |- strategy_II
|              |- prompt_lt.json
|              |- prompt_en.json
|   |- strategy_II
|              |- prompt_appeal_to_fear_lt.json
|              |- ... (files for the remaining techniques in both languages)
|   |- strategy_IV
|              |- prompt_appeal_to_fear_lt.json
|              |- ... (files for the remaining techniques in both languages)
|- scripts/
|   |- output
|   |- gpt.py
|   |- post-procesing.py
|- README.md
```

Fig. 2. The representation of the directory structure. Input prompts are built from templates in prompts/strategy_x with comment content inserted from data/comments.csv. Output files are stored in run specific subfolders of data/strategy_x.

Table 2
Structure of the CSV file of comments.

Field	Type	Description	Example
comment_id	int	The comment id, used to relate comments with emotional manipulation spans.	*3660*
content	string	The comment content (in Lithuanian)	*gal pilietybės klausimą reikia iškelti iš naujo? Jei jie nevykdo Lietuvos į statymų, jei jie viskuo čia nepatenkinti, jeigu jie nelojalūs, tai jie ne piliečiai. Tad visi, kas neturi lenko kortos, tegul laiko visi Lietuvos Konstitucijos egzaminą, lietuvių kalbos egzaminą ir pan., visiems dirbantiems valstybinėje tarnyboje - išlaikyti dar ir ES egzaminą. Kaip yra visame civilizuotame pasaulyje. O kortininkai tegul pasirenka, kur jiems gyventi. Jei buvo klaida padaryta prie Brazausko, reikia ją skubiai atitaisyti.*

Table 3
Comment length statistics.

Number of comments	1000
Average length of comment	238 chars
Longest comment	3413 chars
Shortest comment	16 chars
Per cent of long comments	30,8%
Per cent of short comments	60,9,6%
Per cent of moderate-length comments	8,3%
Technique's spans-per-comment	5,3

Table 4
Statistics of comment article categories.

Articles category	Number of articles
World politics	427
Lithuanian politics	499
Crimes	32
Health	14
Misc	28

Comments are unedited, uncensored, and often use non-standard language. They vary in length, as shown in Table 3. For the percentage distribution analysis, comments were grouped into three length categories by number of characters in the content field: short ($<$ 218 characters), moderate (218–258 characters), and long ($>$ 258 characters). These thresholds were determined empirically based on the distribution of comment lengths in the dataset. The spans-per-comment value represents the average number of annotated emotional manipulation spans per comment (Table 4).

Comments were posted under articles across various categories. Since article categories were not present in the original LITIS corpus, we determined them manually by examining the content of the corresponding articles. By taking into account that our main goal was to identify manipulative language that seeks to influence public opinion on political or electoral issues, most of the articles originate in the political sphere. There are also comments posted on articles from other categories, such as crime and health. Table shows statistics of the distribution of comments by category.

3.2. Gold dataset

We distinguish 14 manipulative techniques (listed in Fig. 1 and Table) used to influence people through emotion. This is a common practice in propaganda and persuasion. The gold dataset is stored in the directory *data/gold_dataset/* in individual files, with each technique assigned its own file (e.g., *gold_dataset/appeal_to_fear.csv, gold_dataset/bandwagon.csv*). The structure of these files as well as relevant examples are shown in Tables 5 and 6.

In Table, we provide statistics on the gold dataset broken down by emotional manipulation techniques. In total, 5300 manipulative spans were annotated across the dataset. For each technique, we show the number of comments in which the technique was identified, the number of spans associated with each technique, and the span length distribution. The number of spans equals the total number of records (rows) in each CSV file in the gold dataset. The number of comments is calculated by counting the unique values in the *comment_id* field of the corresponding CSV files.

The relatively high average number of manipulation techniques per comment exists in the dataset (last row in Table 3), which is primarily attributable to the span-based, overlapping annotation strategy adopted in this work. By permitting nested and intersecting labels, the annotation scheme captures the layered nature of manipulative rhetoric, in which multiple techniques may jointly operate within the same textual fragment rather than appearing as isolated phe-

Table 5
Structure of CSV files of the gold dataset.

Field	Type	Description	Example
gold_id	int	Unique identifier of the gold span.	*885*
comment_id	int	Identifier of the comment containing the span.	*3660*
span	string	The span of the emotional manipulation instance.	*gal pilietybės klausimą reikia iškelti iš naujo?*
start	int	The start character index of the span in the comment. 0-based.	*0*
end	Int	The end character index of the span in the comment. 0-based.	*48*

Table 6
The statistics of the gold dataset.

Technique	Number of Comments	Number of Spans	Length of span, chars		
			Avg.	Min.	Max.
Name-calling	390	617	19	3	175
Labelling	588	945	27	5	175
Flag-waving	198	250	104	12	570
Appeal to fear	170	195	99	9	843
Bandwagon	22	23	85	25	171
Exaggeration	330	501	73	12	390
Minimisation	83	100	112	12	424
Doubt	400	547	97	11	388
Guilt by association	113	125	79	12	239
Reductio ad Hitlerum	40	55	56	8	157
Loaded language	712	1594	53	4	462
Scapegoating	182	210	91	16	301
Slogans	38	45	26	11	56
Support	82	93	70	9	297

nomena. This design choice is consistent with established practices in prior propaganda and manipulation datasets that employ span-level multilabel annotations, e.g., [8,9]. Complementarily, Table provides a technique-level breakdown, thereby contextualising the co-occurrence density observed at the comment level and confirming that it reflects representational fidelity rather than annotation inflation, with clear implications for multilabel and span-aware benchmarking.

The dataset exhibits substantial class imbalance across emotional manipulation techniques (see Fig. 3), with some occurring far more frequently than others, for example, “Loaded language” and “Labelling” techniques have the most examples and “Slogans” and “Bandwagon” are the least common. Similar tendencies with respect to class proportions are also reported in English and other language text datasets [8,9]. This disparity reflects both natural variation in real-world usage and the composition of the selected comment corpus, where certain techniques are underrepresented. Such imbalance should be carefully considered when using the dataset for machine learning, as it can bias models toward dominant classes and degrade performance on rarer techniques. Established strategies for handling class imbalance may therefore be necessary during model training. Additionally, the dataset may serve as a useful resource for augmenting future corpora, particularly in cases where a set of techniques overlap.

3.3. Prompt-generated dataset

During the automated annotation phase, each prompt was subjected to five independent iterations with different temperature parameter values. Therefore, the results directories for each prompt engineering strategy are organised into:

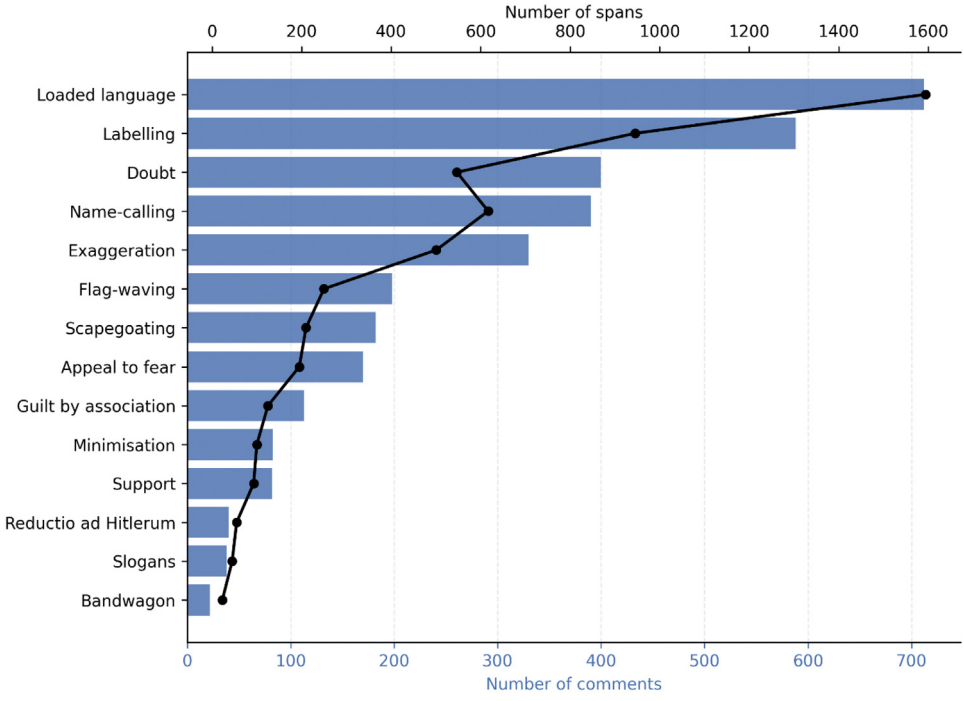

Fig. 3. Per-technique counts of comments (blue bars, bottom axis) and spans (black line, top axis) in the dataset.

- *data/strategy{I-IV}/**run_{1–5}*** – results for each of the 5 runs.
- *data/strategy{I-IV}/run_{1–5}/**temp_{0|1}*** – results, obtained by submitting prompts to the model with different temperature parameters (t = 0 or t = 1).

The dataset contains both raw data returned by the model and pre-processed data with aligned spans and their positions within a comment. Therefore, we further separate the results into the following folders:

- *data/strategy{I-IV}/run_{1–5}/temp_{0|1}/**{post-processed|raw}*** – results for post-processed and raw annotations.

The results of the first and second prompt engineering strategies are stored in *CSV* files named *spans.csv*. The results of the third and fourth strategies are stored in separate files for each emotional manipulation technique (e.g., *appeal_to_fear_spans.csv, bandwagon_spans.csv*, etc.). The file structure of the raw results and the examples is presented in Table 7. Note that the fields vary depending on the strategy used.

The structure of the post-processed results is presented in Table 8. Note that this version of the results has both a span field and its position in a comment.

4. Experimental Design, Materials and Methods

The main processes involved in creating the data source are presented in Fig. 4. During the process, we stored all data in a relational database. The icons in the diagram represent the main table groups used to store comments, golden annotations, prompts, and GPT-4.1 produced annotations (Fig. 5).

Table 7
Structure of CSV files of raw results.

Field	Type	Description	Used for strategies	Example
gpt_results_id	int	Unique identifier of the span.	I, II, III, IV	*1692*
comment_id	int	Identifier of the comment containing the span.	I, II, III, IV	*229,616*
found_ manipulation	string	Whether emotional manipulation is present. Possible values: Taip/Ne (Yes/No).	I, II, III	*Taip*
span	string	Text span returned by the GPT model.	I, II, III, IV	*jeigu jie nelojalūs, tai jie ne piliečiai*
technique	string	Identified manipulation technique. Strategy I outputs are in Lithuanian.	I	*Skirstymas į "mus" ir "juos" (kaltinimas nelojalumu)*
			II	*Labeling*
emotion	string	Emotion intended to be evoked. Strategy I outputs are in Lithuanian.	I	
			II, III	*Negative attitude / Confrontation*

Table 8
Structure of CSV files of post-processed results.

Field	Type	Description	Used for strategies	Example
gpt_results_id	int	Unique identifier of the span.	I, II, III, IV	*1692*
comment_id	int	Identifier of the comment containing the span.	I, II, III, IV	*3660*
found_ manipulation	string	Whether emotional manipulation is present. Possible values: Taip/Ne (Yes/No).	I, II, III	*Taip*
aligned_span	string	Exact span extracted from a comment.	I, II, III, IV	*jeigu jie nelojalūs, tai jie ne piliečiai*
start	int	The start character index of the span in the comment. 0-based.	I, II, III, IV	*117*
end	int	The end character index of the span in the comment. 0-based.	I, II, III, IV	*158*
technique	string	Identified manipulation technique. Strategy I outputs are in Lithuanian.	I	*Skirstymas į "mus" ir "juos" (kaltinimas nelojalumu)*
			II	*Labeling*
emotion	string	Emotion intended to be evoked. Strategy I outputs are in Lithuanian.	I, II, III	*Negative attitude / Confrontation*

At the start of creating the data source, we imported comments from the LITIS text corpus and from the *Facebook* pages of news portals into the database. We then imported 1000 selected comments into Label Studio, an open-source data labelling tool, where annotators manually created annotations to form the gold dataset. These annotations were then transferred to the gold dataset tables. Next, we began generating annotations by prompting GPT-4.1. Researchers initially wrote prompts for all four prompt engineering strategies. We then sent these prompts, along with the comments, to the OpenAI API[1] via a Python script, which inserted the raw results into the GPT annotations tables. We further processed the results using a post-processing script and finally exported data from all tables to the data source files.

[1] https://openai.com/index/openai-api/

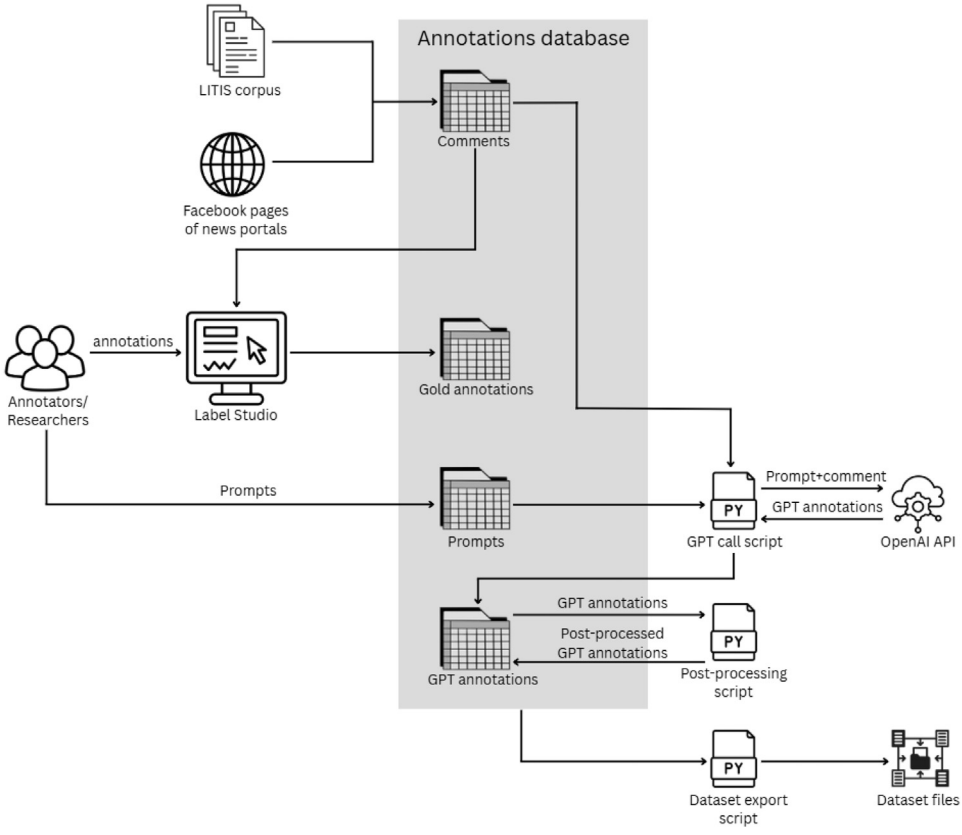

Fig. 4. Workflow of dataset construction.

```
7   # Open AI CLIENT CONFIGURATION
8   client = OpenAI(api_key="WRITE YOUR OWN OPENAI API KEY HERE.")
```

Fig. 5. OpenAI client configuration.

4.1. Annotating the gold dataset

The gold dataset was created in the Label Studio [2] environment by four human annotators, who were expert-level researchers. They manually annotated all 1000 comments following guidelines that provided definitions of 14 techniques and marking rules (see Annex A and B). In drafting the annotation guidelines, we grounded our approach in prior research efforts devoted to constructing annotated datasets for the identification of propaganda and persuasive techniques [8].

Annotators were trained in two stages, (1) first one was comprised of a two-week course, consisting of 2–3 weekly workshops in additions to short consultations. This course tied to the guidelines that were developed based on good practices exemplified by similar studies and provided sufficient theoretical and practical knowledge for initial annotation stage. The second (2) stage took the form of annotating an initial test batch of comments and gathering annotator feedback that subsequently led to the need to review and update current version of guidelines

as ambiguous cases or non-standard practices were identified in order to ensure a unified approach and quality of work offered by all annotators.

Since annotation is often subjective, we used a consensus-based method in which each annotator had to annotate all comments. During annotation, it was first necessary to determine whether the comment contained emotional manipulation. If it did, the annotator could mark the text and select one of the 14 emotional manipulation techniques. Annotations could be added freely, meaning they could vary in length and overlap, or apply to parts of the same fragment multiple times.

We measured the inter-annotator agreement (IAA) calculating *Krippendorff's Alpha_U* [10]. IAA showed moderate reliability 0.59. Agreement varied across techniques, with higher consistency for rhetorical techniques (e.g., Support 0.69, Flag-waving 0.64) and lower agreement for ambiguous techniques (e.g., Reductio ad Hitlerum 0.23, Bandwagon 0.28). A more detailed approach to resolving qualitative disagreements among annotators is defined in the Supplementary Material section, Annex A (Dataset Annotation Guidelines).

Given this moderate level of agreement and the observed variability across techniques, a consensus-building procedure was employed to improve annotation quality and ensure a reliable gold standard. After independently annotating a portion of the comments, the annotators held initial pairwise discussions to coordinate their results. Subsequently, they participated in whole-group discussions to confirm the final annotations and incorporate them into the gold dataset. The annotators then moved on to annotate the next batch of comments, repeating the process.

In order to ensure high-quality data and reduce potential subjectivity during the annotation process, the annotator team employed a set of predefined annotation guidelines that were incrementally updated to address unique cases and standardise the decision-making process. Each guideline update was followed by a new review of previously annotated data. These guidelines provided annotators with formal definitions for each manipulation technique and illustrative examples to ensure mutual understanding of technique manifestations within the comment.

An important part of the annotation process was the support for non-linear and overlapping text segments. In the case of the former one, during early annotation stages, it became apparent that manipulating the technique could be interrupted by an unrelated phrase or even other techniques. In such cases, it was decided that the label should encompass both parts of the technique in question, even if it included additional textual information, as these segments, if separated, would not perform their manipulative function as standalone text. In the case of the latter, overlapping segments were more common; therefore, we recognised that emotional manipulation often operates through layered rhetorical strategies. It was decided to permit the application of multiple labels to the same or intersecting fragments of text as nested annotations, where a broad emotive span (such as an appeal to fear) might contain a more specific, localised technique, such as name-calling. This approach ensured that the dataset could reflect the true complexity of natural language, where a single sentence may simultaneously employ several manipulation techniques.

To ensure that the annotated segments remained semantically coherent even when extracted from the original comment, annotators followed the contextual sufficiency rule. For techniques that manifested in shorter text spans, such as in the case of name-calling, the target of the slur was included within the span whenever possible to provide necessary relational context. Otherwise, the annotation was restricted to the specific pejorative term. Many techniques, however, operated on a broader text span; therefore, these had to be annotated as complete phrases or even as multiple-sentence blocks if they spanned across a larger portion of the comment. This approach ensured that the label in the gold dataset carried enough linguistic information to be interpretable as a standalone instance of manipulation. Additionally, if the same technique was used several times in a sequence, for example, several slurs following each other, all cases were treated as separate instances.

During the annotation process, certain types of ambiguities in comment interpretation were identified and had to be addressed in the guidelines for further work. Common types of ambiguities were as follows:

1. **Accusations vs. Attempts to Scare**. In some cases, annotators had to establish a difference between cases where commenter accuses a person of some negative action or inaction and cases where such comments can actually be interpreted as a deliberate attempt to instil fear as part of Appeal to Fear technique. For example, the comment "*Tomaszewski works for Moscow, not for his constituents"* (original Lithuanian "tamasevskis dirba maskvai o ne jo rinkejams") balances on the verge of being a simple accusation of not caring about his own voters and an attempt to scare the reader that a high-ranking politician may actually work for the interests of an unfriendly foreign power. Due to the geopolitical situation in Lithuania and its emotional impact on local population, it was decided that in this case it falls under manipulation technique rather than semi-neutral statement, even though some arguments can be made for the latter as well.
2. **Limitations of name-calling technique**. While attempts were made to capture cases of negative comparisons and direct name-calling, it had to be acknowledged that popular Lithuanian proverbs and sayings often included mentions of animals. A common one, "Let a pig into the church, and it'll even climb up onto the altar" (original Lithuanian "į leisk kiaulę į bažnyčią, tai ir ant altoriaus užlips"), semantically compares the person in question with a pig, however, pragmatically this is not the case as this proverb means that if you grant a small favour or privilege to a shameless or ungrateful person, they will completely overstep their boundaries and take total advantage of your kindness and is essentially the local equivalent of "give them an inch, and they'll take a mile". In such cases, animal mentions in proverbs were usually excluded from the name-calling technique.
3. **Slogan distinction**. While slogans are usually well-defined and known and serve a distinctive function. In practice, we found that there are slogans that are not established but appear in the form of a slogan or, in some cases, can take the form of a hashtag. These cases had to evaluated individually as they sometimes became trending and were repeated across social media and, therefore, comments for some time, while some were plainly unique. For example, "*Let's vote for Him, and there won't be a war"* (original Lithuanian "balsuokime už Jį , ir karo nebus") may be considered a case of support or a political argument but due to its use in social media campaign as a recurring phrase, it was included in the dataset as a slogan.

Developed guidelines defined the inclusion of punctuation to maintain consistency in the gold dataset. Standard full stops were generally excluded if the technique ended naturally before the end of the sentence. However, in such cases where a specific technique spanned multiple sentences, internal full stops were retained to retain the structural integrity of the original idea. Annotators also intentionally included question marks, exclamation points and ellipses within the span. We considered these elements as intensifiers that emphasised and strengthened the emotional impact of the comment or specific argument.

As an additional supportive mechanism within the annotation process, annotators were asked to assign labels indicating the primary emotion that the author of the comment likely intended to evoke, such as fear or anger. This emotion assignment did not function as an independent analytic layer, but rather as an internal control step designed to guide annotators' decision-making and encourage explicit reflection on whether a comment can reasonably be considered emotionally manipulative.

Finally, this second-tier labelling of emotions identified within individual comments was ultimately excluded from the final dataset due to two primary concerns. (1) The annotation process revealed substantial variability in emotional interpretation across annotators. For example, a statement such as "migrants are thieves and terrorists" could be perceived as evoking fear, distrust, or anger. Due to this inherent subjectivity, emotion annotations differed considerably among the four annotators, resulting in insufficient reliability. (2) While short comments usually represented 1–3 emotions, long ones, spanning 2–3 paragraphs, could eventually contain a whole list of possible emotions, which did not provide much benefit to the final dataset as spans were not labelled in the same way as techniques were, i.e., segments having clear boundaries and more relied on a "feeling".

Table 9
Example of strategy IV prompt for "Doubt" technique identification in English.

System role	**Task:** Identify the passages in the text below that represent the use of the "Doubt" technique. Refer to the definition, prerequisites, and rules of the "Doubt" technique provided. 1. **Definition in English:** A "Doubt" or questioning technique aims to undermine trust or confidence in a person, institution, or claim by raising suspicion, asking leading questions, or suggesting uncertainty. It often relies on rhetorical questions, speculation, or unverified statements rather than factual evidence. 2. A **"doubt"** text fragment must meet all (a and b) **conditions**: (a) The fragment must state or imply that something is incorrect, unreliable, unclear, or questionable. E.g.: "Is this really true?", "Something is wrong here", "This cannot be trusted". (b) The fragment must be aimed at making others doubt a certain position, action, institution, or person. E.g.: "What is really behind this?", "Are they really telling us the truth?" 3. **Rules for forming the answer:** Provide as many lines of answer as there are examples of the \"Doubt\" technique identified in the comment. In each line, provide an accurate, original fragment of the text from the comment, without any corrections or interpretations. The answer must contain only a fragment of text that unambiguously meets both conditions (a and b) of point 2. If the text fragment does not meet at least one condition, the answer should contain the text "DB not found - does not meet the conditions (specify which ones)". The fragment must be continuous, no shorter than a few words, and end with the end of the last word (without punctuation marks).
User template	'{comment}'

In practice, due to these sources of subjectivity, early results from emotion annotation proved unsuitable for further analytical use. Consequently, the research team decided to discontinue this annotation layer and instead focus resources on the continued development and refinement of the manipulation technique dataset.

4.2. Generating annotations by prompting

For all comments, we performed emotional manipulation language detection (i.e., generating annotations) using GPT-4.1 and 30 distinct prompt templates developed with four prompt engineering strategies: one prompt for Strategy I, one for Strategy II, and 14 prompts each for Strategies III and IV, corresponding to individual emotional manipulation techniques. All prompts are available in the dataset's top-level *prompts* directory, and their English translations are provided in Annex C.

Prompt structure and level of detail depend on the strategy, where strategy I and II prompts are more general and not tailored to each technique and strategy III and IV are technique-specific with definition of technique, conditions and rules. Example of prompt for "Doubt" technique identification (original file – *prompts/strategy_IV/prompt_doubt_en.json*) is provided in Table 9.

To make it easier to reproduce our experiments on emotional manipulation language identification, we share the Python scripts used to identify manipulative language spans in comments with the GPT-4.1 model and to post-process the results. We want to highlight that the Python scripts shown in Fig. 5 perform the same functions as the files provided in the data source. The only difference is that the scripts in the data source are adjusted to work with text files.

During our experiments, we observed that the model did not always return the exact excerpts of emotional manipulation instances from the comments, even though we requested this in the instructions. It sometimes altered the morphological form of phrases, replaced pronouns, punctuation marks, and so on. Therefore, we post-processed the raw results and aligned GPT-identified annotations with the comments. We used a fuzzy substring-matching algorithm designed to identify approximate occurrences of a query fragment within a longer text, ensuring that matches correspond to complete words. The algorithm uses a sliding-window comparison

```
# OpenAI API CALL CONFIGURATION
# Prompt engineering strategy. Allowed values: I, II, III or IV
strategy = "II"

# Language setting that will be used to determine the prompt file name. Allowed values: lt or en.
# If language is set to lt, prompt file name will be prompt_lt.json. If language is set to en,
# prompt file name will be prompt_en.json.
language = "lt"

# for I and II strategy, technique is not specified, so it will be None. For III and IV strategies,
# technique is specified, so it will be one of the techniques from the techniques dictionary.
technique = techniques["nc"]

# GPT model.
model = "gpt-4.1"

# GPT model temperature. Typically, ranges from 0 to 1. Higher values like 0.8 will make the output
# more creative, while lower values like 0.2 will make it more focused and deterministic.
temp = 0

# GPT run number. This will be used to determine the results file name and path.
run = "1"
```

Fig. 6. The experiment's parameters.

and returns the most similar fragment from the original comment. In Annex D, we report an accuracy assessment conducted on a sampled subset of the data, using a manually inspected validation set, and present quantitative results for character offset correctness. In addition, we include illustrative examples that demonstrate both the strengths and the known limitations of the algorithm under different alignment conditions.

Another thing we noticed is that GPT-4.1 tends to produce inconsistent results when run multiple times. Each prompt was subjected to five independent iterations with temperature set at t = 1 to enable maximum "creativity" of the model. Interestingly, during preliminary experiments, we noticed that GPT produced slight variations in responses to the same prompts across different runs, even with the temperature set at t = 0. Although it should produce no variability at t = 0, we speculate that this is not the case due to some design decision by OpenAI. Therefore, we found it necessary to also use five iterations for the prompts when running with the temperature set at t = 0.

Further, we present instructions for reproducing our experiment and creating emotionally manipulative language annotations using GPT-4.1 via the OpenAI Completions API. It can be done by calling the *gpt.py* script. This script performs emotional language identification for all 1000 comments. The script automatically reads comments from *comments.csv* and the required prompt from *prompts*, submits the prompt to GPT, and writes both raw and post-processed results to output files. The script calls the OpenAI Completions API for each comment. By default, the results are written to the *scripts/output* directory, using the directory structure presented in Fig. 2.

The setting to change the OpenAI API client key is presented in Fig. 5.

The experiment's parameters are shown in Fig. 6.

Depending on the selected prompt engineering strategy (I, II, III, or IV) and language (lt or en), the corresponding prompt will be retrieved from the prompts catalogue (e.g., *prompts/strategy_I/prompt_lt.json, prompts/strategy_II/prompt_en.json*). The prompt files store the templates, used for interactions with the OpenAI Completion API. The files contain two main fields: system, which defines the *system* instructions for the annotation task and the expected output format, and *user_template*, which contains the user message template. During processing, the user template is populated with the comment text to be analysed and submitted to the model together with the system instructions. The structure of these JSON files is presented in Table 10.

Table 10
Structure of JSON files of prompts.

Field	Type	Description	Example
prompts_id	int	Unique identifier of the prompt.	33
system	string	Defines the prompt with the expected output format.	Comments will be provided below. Please evaluate whether they are emotionally manipulative (aiming to evoke emotions in the reader). For each comment, provide as many as there are recognized examples of manipulation. The response structure should match the following template: < manipulative or not (\"Yes\"/\"No\") > | ⟨ manipulation technique ⟩ | example from a comment | ⟨ intended to evoke emotion ⟩ |. The example of the technique's use must be an accurate, unedited excerpt from the comment.
user_template	string	Contains the template for the user message. Only the comment is sent as a user message.	{comment}

The strategy parameter is related to the technique parameter, which must be set when strategy III or IV is chosen. The prompts for these strategies are designed to identify instances of emotional manipulation using a specific technique. This parameter must be set to one of the keys (i.e., the abbreviation of a technique) in the techniques dictionary. The abbreviations are used to shorten names. For example, if one wants to use the "name_calling" technique, this parameter must be set to techniques["nc"]. Depending on the selected parameter, the corresponding prompt file for that technique will be used (e.g., *prompts/strategy_III/prompt_appeal_to_fear_en.json* and *prompts/strategy_III/prompt_bandwagon_en.json*). The technique parameter is ignored when strategy I or II is selected.

There are two parameters related to the GPT model. The model allows choosing which model to use. The model's temperature parameter, temp, yields more deterministic responses at lower values (around 0) and less deterministic responses at higher values (around 1).

The run parameter allows the same experiment to be repeated several times to assess the stability of the model. The results will be saved in run-specific directories (e.g., *output/strategy_I/run_1*).

The script can be executed from the command line:

python gpt.py

To run the script, Python version 3.10 or higher must be installed. The experiments were conducted using the following libraries:

- openai (OpenAI Python SDK) version 2.26.0
- rapidfuzz version 3.14.3
- pandas version 2.3.3

Newer versions of these libraries may also work, but they have not been tested. The exact versions used in the experiments can be installed with pip by running the following commands:

python -m pip install openai==2.26.0

python -m pip install rapidfuzz==3.14.3

python -m pip install pandas==2.3.3

If permission errors occur on Unix-based systems, ensure the output directory is writable, for example:

chmod -R u+rwx output

Limitations

There is a relatively small number of comments/examples representing some techniques, i.e., *Bandwagon, Slogans*, and *Reductio ad Hitlerum*. Instances of these techniques were rare in the source media of the dataset.

The dataset reflects communication patterns typical of users who actively participate in online news portals and public social media discussions and may not capture the full demographic diversity of the Lithuanian-speaking population. As a result, the linguistic style and prevalence of emotional manipulation techniques observed in the dataset may differ from those found in broader societal discourse.

The selection of source material may also introduce political and topical bias. Because the dataset includes comments associated with news and public discussions, it contains a higher proportion of content related to socially and politically salient issues. These contexts are known to elicit stronger emotional responses and more frequent use of manipulative rhetoric, which may skew the distribution of techniques toward those commonly used in contentious or polarizing debates, while more neutral or subtle forms of persuasion remain underrepresented.

Additionally, the use of keyword-based retrieval to identify candidate comments introduces a systematic bias toward texts that explicitly contain markers of emotionally charged or manipulative language. While this approach improves coverage of targeted phenomena, it may still favour more explicit instances of manipulation and underrepresent subtler forms that do not rely on identifiable lexical cues. Consequently, the dataset should be understood as a targeted resource for studying emotional manipulation rather than a statistically representative sample of all Lithuanian online discourse.

Ethics Statement

We have read and followed the ethical requirements for publication in Data in Brief. We confirm that the current work does not involve human subjects or animal experiments. We ensure that the data utilised in this work was obtained from the appropriate primary sources and reused with their consent. The authors have complied with ethical standards and, when necessary, have acquired permission to reuse the data.

Data Availability

Lithuanian Emotional Manipulation Dataset (Original data) (Zenodo).

CRediT Author Statement

Rita Butkienė: Conceptualization, Data curation, Methodology, Investigation, Writing – original draft; **Algirdas Šukys:** Data curation, Investigation, Software, Writing – original draft; **Edgaras Dambrauskas:** Data curation, Resources, Investigation; **Voldemaras Žitkus:** Data curation, Investigation, Software; **Linas Ablonskis:** Validation, Writing – review & editing; **Evaldas Vaičiukynas:** Visualization; **Paulius Danėnas:** Software; **Rimantas Butleris:** Supervision, Resources, Project administration, Funding acquisition.

Acknowledgements

Funding: This work was conducted as part of the execution of the project “Mission-driven Implementation of Science and Innovation Programs” (No. 02-002-P-0001), funded by the Economic Revitalisation and Resilience Enhancement Plan “New Generation Lithuania.”

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2026.113000.

References

[1] D. Amilevičius, M. Petkevičius, LITIS v.1. Vytautas Magnus University. CLARIN-LT Repository, 2016 https://hdl.handle.net/20.500.11821/11.
[2] T. Maxim, M. Malyuk, A. Holmanyuk, N. Liubimov, Label studio: Data labeling software, 2025. https://github.com/HumanSignal/label-studio.
[3] K. Hamilton, L. Longo, B. Bozic, GPT assisted annotation of rhetorical and linguistic features for interpretable propaganda technique detection in news text. (2024) 1431–1440. https://doi.org/10.1145/3589335.3651909.
[4] A. Sahitaj, P. Sahitaj, V. Solopova, J. Li, S. Möller, V. Schmitt, Hybrid annotation for propaganda detection: integrating LLM pre-annotations with human intelligence, (2025) 215–228. https://doi.org/10.18653/v1/2025.nlp4pi-1.18.
[5] A. Gaeta, V. Loia, A. Lorusso, F. Orciuoli, A. Pascuzzo, Towards a LLM-based intelligent system for detecting propaganda within textual content, Comput. Electr. Eng. 128 (Part B) (2025) 110765, doi:10.1016/j.compeleceng.2025.110765.
[6] I. Rizgelienė, V. Zubaitienė, N. Maliukevičius, V. Marcinkevičius, HALT-PROP: human-annotated Lithuanian textual corpus for propaganda narratives and techniques, Sci. Data 13 (2025) 47, doi:10.1038/s41597-025-06367-w.
[7] R. Butkienė, A. Šukys, E. Dambrauskas, V. Žitkus, L. Ablonskis, E. Vaičiukynas, P. Danėnas, R. Butleris, Lithuanian emotional manipulation dataset [Data set], 2026. https://zenodo.org/records/18890147.
[8] G. Da San Martino, S. Yu, A. Barrón-Cedeno, R. Petrov, P. Nakov, Fine-grained analysis of propaganda in news articles, (2019) 5636–5646. https://doi.org/10.18653/v1/D19-1565.
[9] J. Piskorski, D. Dimitrov, F. Dobranić, M. Ernst, J. Haneczok, I. Koychev, N. Ljubešić, M. Marcińczuk, A. Modzelewski, I. Moravski, SlavicNLP 2025 shared task: detection and classification of persuasion techniques in parliamentary debates and social media, (2025) 254–275. https://doi.org/10.18653/v1/2025.bsnlp-1.27.
[10] K. Krippendorff, Content Analysis: An Introduction to Its Methodology, Sage publications, 2018.