Method of semantic features estimation for political propaganda techniques detection using transformer neural networks

Iurii Krak[1,2], Maryna Molchanova[3], Volodymyr Didur[3], Olena Sobko[3], Olexander Mazurets[3] and Olexander Barmak[3]

[1]*Taras Shevchenko National University of Kyiv, 64/13 Volodymyrska Str., Kyiv, 01601, Ukraine*
[2]*Glushkov Institute of Cybernetics of NAS of Ukraine, 40 Glushkov Ave., Kyiv, 03187, Ukraine*
[3]*Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, 29016, Ukraine*

Abstract

The paper is devoted to creation and approbation of method of semantic features estimation for political propaganda techniques detection using transformer neural networks was proposed, which allows to achieve an increase in explainability level of decisions made by transformer neural networks regarding the presence of political propaganda techniques by assessing semantic features manifestation, as well as to increase the accuracy of identification of gray and black propaganda techniques. This effect is achieved by using additional semantic features for political propaganda techniques and modified architecture of transformer neural networks that analyze not only text data, but also additional vector of numerical values of semantic features. Applied study of developed method of semantic features estimation for political propaganda techniques detection using transformer neural networks revealed significant increase in the detection accuracy of 5 political propaganda techniques ("Appeal to Fear-Prejudice", "Repetition", "Causal Oversimplification", "Minimisation" and "Appeal to Authority") and slight increase in the detection accuracy of 4 political propaganda techniques ("Red Herring", "Exaggeration", "Whataboutism" and "Thought Terminating Cliches"), the method was recognized as more effective for these techniques than existing analogues. For 8 political propaganda techniques ("Loaded Language", "Labeling", "Name Calling", "Black and White Fallacy", "Slogans", "Doubt", "Reductio ad Hitlerum" and "Flag-Waving") detection accuracy increased slightly or did not increase, the method was recognized as parity for these techniques with existing analogues. For developed method, minimum accuracy of propaganda techniques detection is 0.85, maximum accuracy is 0.97, and average accuracy is 0.89. The developed method has significant potential for achieving the Sustainable Development Goals No. 4 and No. 16, in particular in increasing media literacy and critical thinking among the population. This will contribute to strengthening democratic institutions and ensuring transparency in political decision-making, which is important step in fight against disinformation and manipulation.

Keywords

propaganda semantic features, political propaganda techniques, transformer neural networks, NLP

1. Introduction

In the modern information society, political propaganda has become a powerful tool for manipulating public consciousness [1]. The use of propaganda techniques in media and social networks contributes to distortion of reality, spread of disinformation and reduction of trust in information sources. This negatively affects civil society and democracy, creating distorted perception of events [2], as well as influencing elections and political decisions. The development and spread of political propaganda is also facilitated by the development of NLP, which generates large volumes of artificially created texts, which are increasingly difficult to distinguish from those written by person [3].

At same time, existing methods for detecting propaganda remain insufficiently effective, as they do not cover the entire spectrum of manipulations and do not take into account additional semantic

CS&SE@SW 2024: 7th Workshop for Young Scientists in Computer Science & Software Engineering, December 27, 2024, Kryvyi Rih, Ukraine

yuri.krak@gmail.com (I. Krak); m.o.molchanova@gmail.com (M. Molchanova); pravetz@ukr.net (V. Didur); olenasobko.ua@gmail.com (O. Sobko); exe.chong@gmail.com (O. Mazurets); alexander.barmak@gmail.com (O. Barmak)
0000-0002-8043-0785 (I. Krak); 0000-0001-9810-936X (M. Molchanova); 0009-0008-2279-1487 (V. Didur); 0000-0001-5371-5788 (O. Sobko); 0000-0002-8900-0650 (O. Mazurets); 0000-0003-0739-9678 (O. Barmak)

features that are characteristic of propaganda.

The methodology for evaluating semantic features for detecting political propaganda techniques using neural networks-transformers contributes to the achievement of Sustainable Development Goal (SDG) No. 16 (Peace, justice and strong institutions) by ensuring transparency of the information environment and increasing trust in institutions, as well as SDG No. 4 (Quality education) by developing media literacy and critical thinking of the population, which allows for effective counteraction to disinformation [4]. This emphasizes the importance of developing automated tools for detecting propaganda manipulations, which will contribute to a more conscious perception of information by users.

The paper aim is to increase the level of decisions explainability made by neural networks-transformers regarding the detection of used political propaganda techniques and increase the accuracy of identifying gray and black propaganda techniques.

The main paper contribution is proposed method of semantic features estimation for political propaganda techniques detection using transformer neural networks, which differs from existing ones by additional explanation of decisions made regarding the presence of used political propaganda techniques based on set of available semantic features and modified architecture of transformer neural networks that analyze not only text data, but also additional vector of features numerical values.

2. Related works

The problem of detecting propaganda remains relevant, as new methods of influencing the audience are constantly emerging, and techniques for synthetic text generation for the dissemination of propaganda messages are also developing. In this regard, an important task is the continuous monitoring of modern approaches to the creation of such content and the improvement of tools for their identification, which will contribute to increasing the level of information security and effective counteraction to disinformation. Therefore, the task of detecting political propaganda is in the field of view of scientists in the field of NLP.

Research [5] shows mixed results regarding the ability of LLMs to identify propaganda techniques in news texts. 14 techniques were considered: “Appeal to Authority”, “Appeal to fear prejudice”, “Bandwagon, Reductio ad hitlerum”, “Black-and-White Fallacy”, “Causal Oversimplification”, “Exaggeration, Minimisation”, “Doubt”, “Flag-Waving”, “Loaded Language”, “Name Calling, Labeling”, “Repetition”, “Slogans”, “Thought-terminating Cliches”, “Whataboutism, Straw Men, Red Herring”. In experiments conducted on the annotated SemEval2020 Task 11 corpora, the best performance reached 64.53% for the recall metric and 81.82% for the precision metric, but no model surpassed the baseline F1 score of about 50%. The maximum F1 score of 20% was significantly lower than the baseline, highlighting the limitations of generative models in reproducibility. When analyzing Polish texts, GPT-4 showed an accuracy of 74% for binary classification (detecting the presence of propaganda) and 69% for the classification of specific techniques. These results indicate the relative success of the model in low-resource conditions, although its performance remains far from the level required for practical use.

The use of machine learning to detect propaganda content on social media was also investigated by Khanday et al. [6]. Given that platforms such as Twitter are often used to manipulate user behavior, the authors emphasize the need for automated approaches to identifying such content. Using data collected using the social network API, the effectiveness of various models was evaluated. The results showed that neural networks, in particular the Long Short Term Memory (LSTM) architecture, demonstrate high accuracy in detecting propaganda. The achieved accuracy level was 77.15%, which indicates the promise of this approach. The authors note that the use of more modern models such as BERT can provide even better results in future studies.

Chaudhari and Pawar [7] explores the problem of identifying propaganda in media in resource-constrained languages, in particular Hindi, which is complicated by the lack of necessary linguistic tools and pre-trained models. Despite significant progress in detecting propaganda for English-language content, classifying propaganda in Hindi requires the creation of new approaches and localized resources. The study developed the H-Prop-News corpus to create vector representations of words (Word2vec),

which became the basis for training the models. Three deep neural networks (CNN, LSTM, Bi-LSTM) and four transformer models (multi-lingual BERT, Distil-BERT, Hindi-BERT, Hindi-TPU-Electra) were tested. The best results were obtained by multi-lingual BERT and Hindi-BERT, which demonstrated the highest F1 score of 84% on the test dataset. The researchers used the RoBERTa language model to detect propaganda techniques in news articles [8]. The model was evaluated using the SemEval-2020 Task 11 reference dataset, showing the ability to detect complex propaganda techniques, outperforming the baseline model with an F1-score of 60.2%. In turn, another study by Jones [9] analyzed the capabilities of large language models (LLMs), in particular the GPT-3.5-Turbo model from OpenAI, to detect signs of propaganda in news. Using the technology underlying ChatGPT, the researchers examined texts for the presence of propaganda techniques identified in previous works [10]. A refined query was developed, which was combined with news from the Russia Today (RT) network and the SemEval-2020 Task 11 dataset to identify propaganda techniques. The results showed that the LLM technology is able to make reasonable conclusions about propaganda, although the detection accuracy is only 25.12% on the SemEval-2022 dataset. However, it has the potential to become a useful tool for detecting propaganda among media consumers and journalists.

In [11], the performance of BERT, RoBERTa, and DeBERTa models combined with different data augmentation methods for detecting propaganda texts is analyzed. Abdullah et al. [11] achieved an F1 micro of 60% on the test set using an ensemble of these models. Another interesting direction is the study of the language of propaganda and its stylistic features [12]. In [13], the PPN (Propagandist Pseudo-News) dataset is presented, which is a multimodal, multilingual dataset consisting of news articles extracted from websites recognized as sources of propaganda by expert agencies. A limited sample from this dataset was randomly mixed with materials from the French press, and their URLs were masked to conduct an annotation experiment in which people used 11 different labels. The results showed that annotators were able to reliably distinguish between the two types of press for each of the labels. The paper also proposes different NLP methods to detect features used by annotators and compare them with machine classification. For this purpose, tools such as VAGO to measure the fuzziness and subjectivity of discourse, TF-IDF as a base model, and four different classifiers are used: two models based on RoBERTa, CATS using syntax, and one XGBoost combining syntactic and semantic features.

The study by Krak et al. [14] considers the detection of 17 propaganda techniques: “Appeal to fear-prejudice”, “Causal Oversimplification”, “Doubt”, “Exaggeration”, “Flag-Waving”, “Labeling”, “Loaded Language”, “Minimisation”, “Name Calling”, “Red Herring”, “Repetition”, “Appeal to Authority”, “Black and White Fallacy”, “Reductio ad hitlerum”, “Slogans”, “Thought terminating Cliches”, “Whataboutism”. Higher efficiency of using neural network models of transformers compared to recurrent models and traditional machine learning approaches was noted. At the same time, insufficient focus on the detection of propaganda techniques by semantic features negatively affected the detection accuracy indicators.

Thus, from the reviewed publications it is summarized that despite the high popularity and importance of this scientific direction, there is still a need to create methods that can increase the accuracy of detecting both propaganda in general and its individual techniques. It is noted that the detection of semantic features can increase the accuracy of detecting political propaganda techniques. Also, the low explainability of neural network solutions is a problem.

3. Method design

Method of semantic features estimation for political propaganda techniques detection using transformer neural networks is intended for evaluating text content for the presence of political propaganda techniques and determining strength of their manifestations. Additional semantic features [15] mean various text features characteristic of certain propaganda techniques. Table 1 shows the averaged indicators of the strength of semantic features manifestations, such as “Emotionality of the text” [16], “Bullying” [17], “Fear” [18] and “Hate speech” [19] in the context of political propaganda techniques. Table 1 is derived by analyzing reference texts containing propaganda techniques.

Table 1
Average strength of semantic features manifestations for propaganda techniques.

Propaganda techniques	Text emotionality	Bullying	Fear	Hate speech
"Appeal to Fear-Prejudice"	0.7	0.5	0.8	0.7
"Causal Oversimplification"	0.4	0.4	0.4	0.3
"Doubt"	0.5	0.4	0.7	0.5
"Exaggeration"	0.8	0.4	0.7	0.6
"Labeling"	0.7	0.4	0.5	0.7
"Loaded Language"	0.8	0.4	0.5	0.8
"Minimization"	0.4	0.4	0.4	0.3
"Name Calling"	0.8	0.7	0.5	0.8
"Reductio ad Hitlerum"	0.9	0.7	0.9	0.9
"Whataboutism"	0.5	0.4	0.4	0.7

To estimate the strength of semantic features, the set of neural networks of transformer architecture was used, namely RoBERTa [20], pre-trained on corresponding binary datasets taken from Kaggle. The datasets used, which are originally in English, were translated into Ukrainian using machine translation before use. The RoBERTa version was chosen because it is capable of working with Ukrainian language and is an optimized version of BERT [21].

The scheme of method of semantic features estimation for political propaganda techniques detection using transformer neural networks in text content, which is aimed at increasing the accuracy of detecting political propaganda techniques, is shown in figure 1.

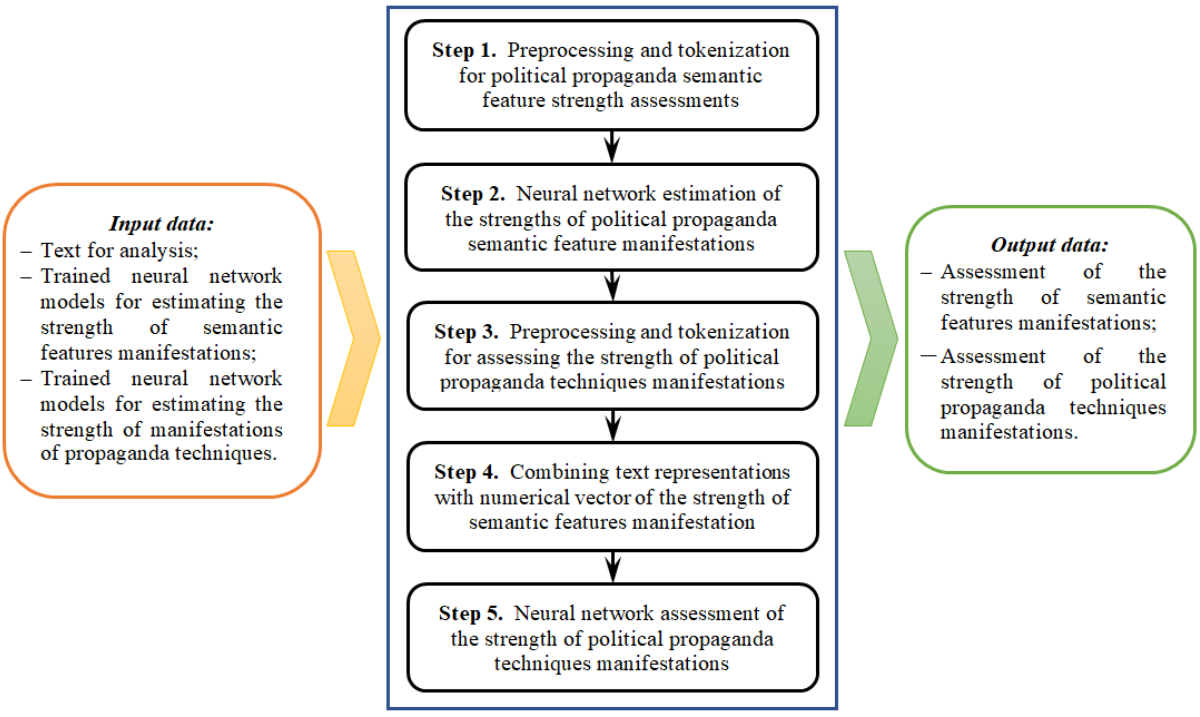

Figure 1: Scheme of method of semantic features estimation for political propaganda techniques detection using transformer neural networks.

The input data of the method are text for analysis, trained neural network models for assessing the strength of semantic features manifestations, and trained transformer neural network models for assessing the strength of political propaganda techniques manifestations.

The Step 1 is preprocessing and tokenization for assessing the strength of political propaganda semantic features manifestations. It includes the removal of stop words and tokenization for each of

the 4 neural networks that determine “Text Emotionality”, “Bullying”, “Fear” and “Hate Speech”. The removal of stop words is common to all neural networks, but tokenization is adaptive for each of neural networks by the tokenizer used during training.

In Step 2, neural network estimates the strength of semantic features manifestations. The score is displayed in the range from 0 to 1, where 0 is the absence of semantic feature manifestation, and 1 is the undeniable presence.

Step 3 involves preprocessing and tokenization to estimate the strength of political propaganda techniques used. It also includes tokenization adapted for each of 17 trained transformer neural networks.

In the Step 4 of combining text representations with numerical vector of the strength of semantic features, different types of data are integrated to improve the classification of the political propaganda techniques used. The data transformation process in the process of neural network detection of semantic features and political propaganda techniques is shown in figure 2.

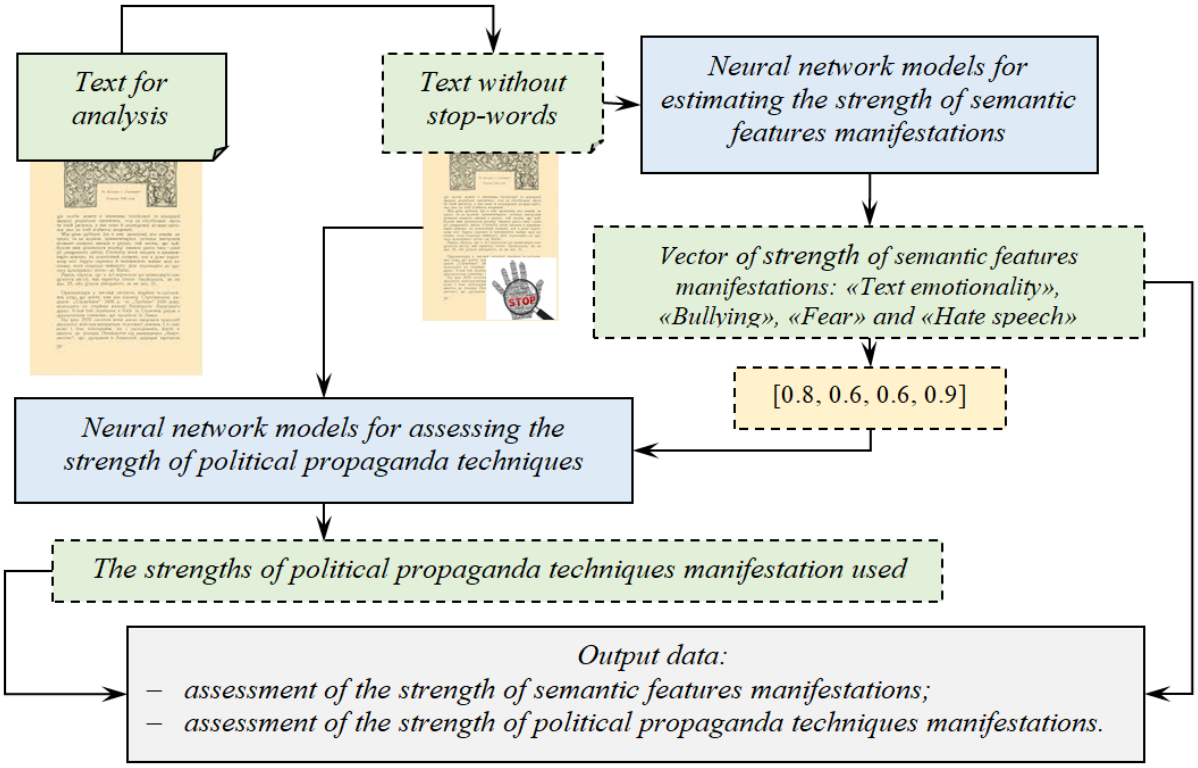

Figure 2: Data transformation scheme for classifying propaganda techniques by semantic features.

Neural network estimation of the strength of political propaganda techniques is the last Step 5, which is performed on the merged data by modified transformer neural network architecture. Method of semantic features estimation for political propaganda techniques detection using transformer neural networks consists in using deep learning architectures, in particular BERT-based models, supplemented with additional numerical vectors that express the strength of semantic features manifestation. The main idea of this is to improve the accuracy of classification of texts containing propaganda techniques by combining contextual information from the text and specific features that indicate various emotional or manipulative elements. The result in the form of a neural network estimation of the available political propaganda techniques will have a partial additional explanation in the form of neural network estimates of the presence of semantic features.

The transformer model architecture includes two main components: the BERT model, which is used to obtain contextual vector representations of the text, and an additional layer for processing numerical semantic features. The BERT model efficiently encodes text, taking into account both the previous and the following words, which allows for a deeper understanding of semantic connections. At the same

time, numerical semantic features pass through separate layer, which highlights important features and reduces their dimensionality.

After that, the outputs from both components are combined and passed to the final classifier, which determines the probability of propaganda techniques presence in the text. This approach allows the model to take into account both deep semantic connections and specific emotional and manipulative features, which significantly improves the accuracy and reliability of detecting political propaganda techniques.

4. Experiment

To test the method, software was created in the form of web application for classifying political propaganda techniques by semantic features (figure 3). The following tools were used to develop the web application: Flask microframework [22] and Python programming language. The capabilities of Google Colab cloud service were used to train neural network models [23]. Used libraries for working with data and neural networks: Sklearn [24], Tensorflow [25], NumPy [26], Pandas [27].

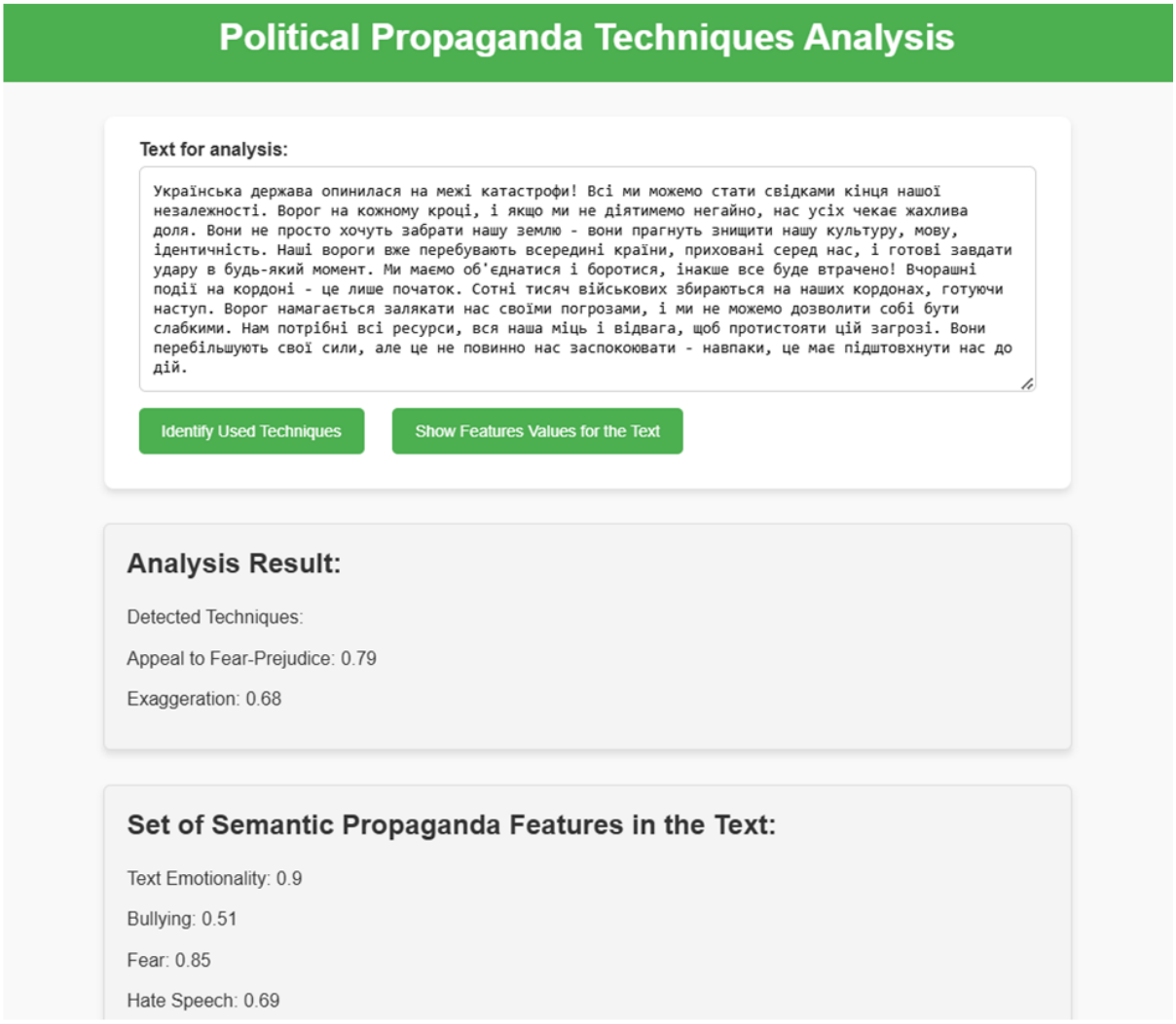

Figure 3: View of developed web application for classifying political propaganda techniques.

To train the transformer models, the dataset from previous studies was used, which was modified so that the text containing each political propaganda technique was placed in separate directory. After such redistribution, statistics were derived for the available texts representing political propaganda techniques. The distribution statistics are shown in figure 4.

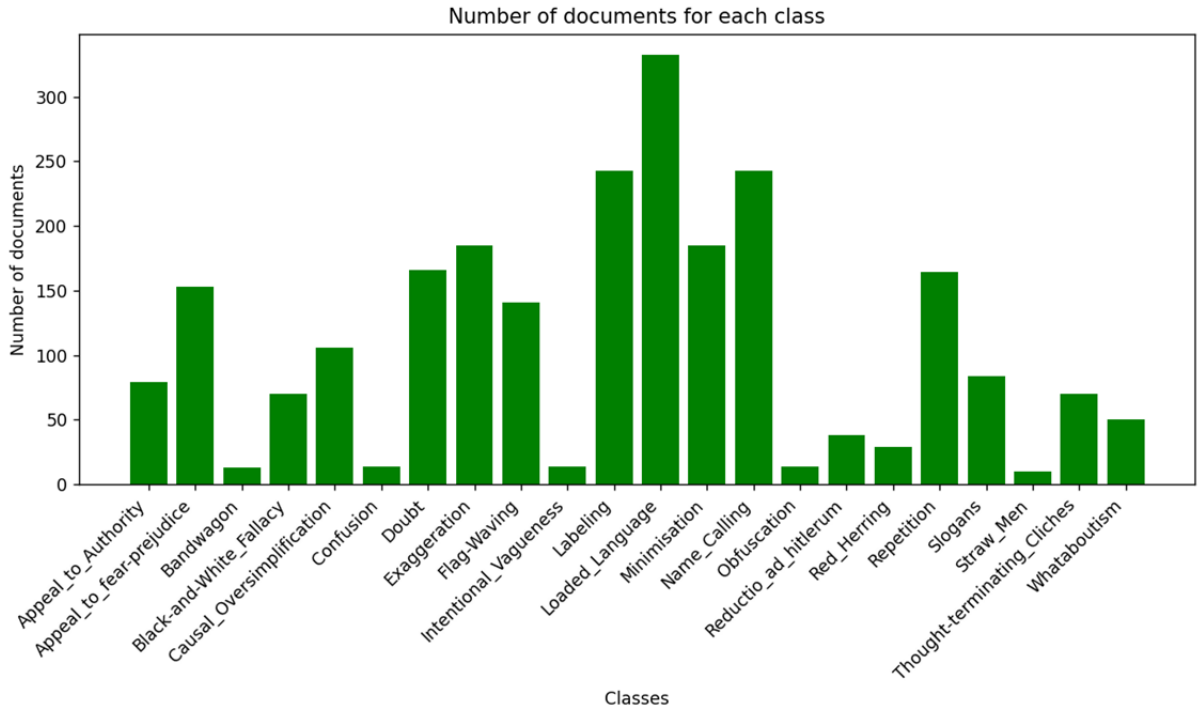

Figure 4: Statistics of the number of texts representing propaganda techniques, pcs.

As can be seen from figure 4, some political propaganda techniques, such as “Bandwagon”, “Confusion”, “Intentional Vagueness”, “Obfuscation” and “Straw Men”, are listed in a critically low number, so creating separate trained neural networks-transformers for them is not advisable. These data are combined into the category “Other propaganda techniques”, but in such a way that the available set does not contain other political propaganda techniques other than the five listed. Since the neural networks-transformers are already pre-trained, further training can be applied to the remaining political propaganda techniques. From the dataset considered above, for each of 17 typical transformer models, child datasets were formed that satisfy the following requirements [14]:

- have texts with specific propaganda technique;
- as opposed to using the set “Other Propaganda Techniques” supplemented with texts without propaganda and texts representing other propaganda techniques different from the target type.

These datasets undergo additional labeling by neural networks before training to identify semantic numerical features and are subsequently trained for each political propaganda technique.

5. Results and discussion

The results of experiments using the developed software implementation for classifying propaganda techniques by semantic features are given in table 2.

Graphical representation of table 2 results is shown in the diagram in figure 5.

The proposed method of semantic features estimation for political propaganda techniques detection using transformer neural networks to increase the accuracy of detecting political propaganda techniques demonstrates noticeable improvements in accuracy compared to well-known analogues [10] and our own previous studies [14]. The overall analysis of the results indicates a significant increase in the effectiveness of the new method, which indicates its potential for use in automated media content analysis systems.

A slight improvement in accuracy (by 0.01) is observed for several political propaganda techniques: “Appeal to fear-prejudice”, “Causal Oversimplification”, “Minimisation”, “Repetition”, “Appeal to Authority”. The “Appeal to fear-prejudice” technique showed a slight increase from 0.87 to 0.88, “Causal

Table 2
Comparison of accuracy of proposed and existing approaches.

Political propaganda techniques	Existing analogues	Previous research	Developed method
“Appeal to fear-prejudice”	0.77	0.87	0.88 (+0.01)
“Causal Oversimplification”	0.7	0.82	0.83 (+0.01)
“Doubt”	0.17	0.93	0.91 (-0.02)
“Exaggeration”	0.54	0.8	0.82 (+0.02)
“Flag-Waving”	0.64	0.92	0.91 (-0.01)
“Labeling”	0.47	0.96	0.93 (-0.03)
“Loaded Language”	0.54	0.97	0.97 (0.00)
“Minimisation”	-	0.9	0.91 (+0.01)
“Name Calling”	0.47	0.92	0.9 (-0.02)
“Repetition”	0.36	0.94	0.95 (+0.01)
“Appeal to Authority”	0.77	0.89	0.9 (+0.01)
“Black and White Fallacy”	0.54	0.91	0.91 (0.00)
“Reductio ad hitlerum”	0.25	0.87	0.87 (0.00)
“Red Herring”	0.39	0.8	0.89 (+0.09)
“Slogans”	0.76	0.86	0.85 (-0.01)
“Thought terminating Cliches”	0.53	0.8	0.83 (+0.03)
“Whataboutism”	0.39	0.79	0.83 (+0.04)

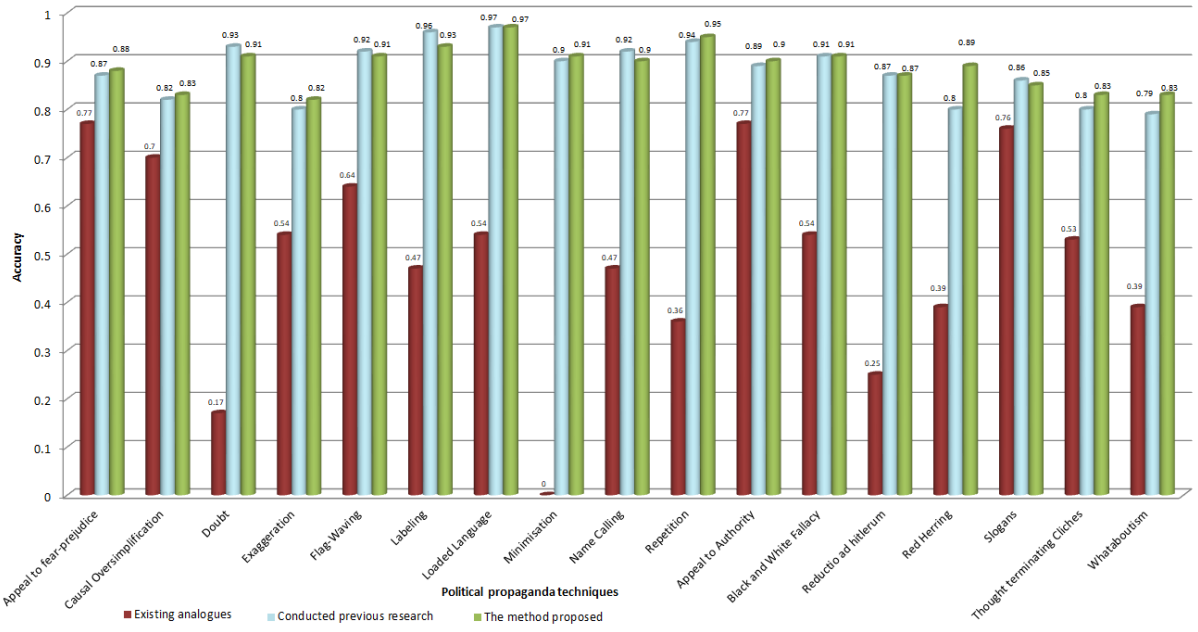

Figure 5: Comparison of accuracy of detecting political propaganda techniques.

Oversimplification” increased from 0.82 to 0.83, “Minimisation” showed an increase from 0.90 to 0.91, and “Repetition” from 0.94 to 0.95. “Appeal to Authority” also showed an increase in accuracy from 0.89 to 0.90. These improvements indicate the effectiveness of the new method in increasing the accuracy of detecting these political propaganda techniques. Several political propaganda techniques showed noticeable improvements. The “Exaggeration” technique improved from 0.80 to 0.82, the “Red Herring” technique showed a significant improvement from 0.80 to 0.89, indicating success in detecting this technique (an increase in accuracy of 0.09). The “Whataboutism” technique also showed significant improvement from 0.79 to 0.83, and the political propaganda technique “Thought terminating Cliches” has an improvement from 0.80 to 0.83.

There is a decrease in accuracy for some techniques. For example, the propaganda technique

"Labeling" decreased from 0.96 to 0.93, and the technique "Name Calling" from 0.92 to 0.90. The accuracy of the technique "Slogans" decreased from 0.86 to 0.85, and the technique "Doubt" from 0.93 to 0.91. This indicates the need for further optimization of the methods for detecting these specific political propaganda techniques.

There are also a number of techniques that have maintained their accuracy unchanged. "Loaded Language" remained at 0.97, "Black and White Fallacy" at 0.91, and "Reductio ad hitlerum" at 0.87. This indicates the stability of the proposed method in detecting these techniques.

In addition to increasing accuracy, proposed method also allows for additional clarification by comparing results obtained with table 1. Accordingly, when analyzing the example from figure 3, within the text "*The Ukrainian state is on the brink of disaster! We could all witness the end of our independence. The enemy is at every turn, and if we do not act immediately, a terrible fate awaits us all. They do not just want to take our land – they seek to destroy our culture, language, identity. Our enemies are already inside the country, hidden among us, and ready to strike at any moment. We must unite and fight, otherwise all will be lost! Yesterday's events on the border are just the beginning. Hundreds of thousands of soldiers are gathering on our borders, preparing an offensive. The enemy is trying to intimidate us with his threats, and we cannot afford to be weak. We need all the resources, all our strength and courage to confront this threat. They exaggerate their strength, but this should not calm us – on the contrary, it should push us to action.*" (translated from Ukrainian), use of the propaganda techniques "Appeal to Fear-Prejudice" with a score of 0.79 and "Exaggeration" with a score of 0.68 was determined. These political propaganda techniques, according to Table 1, are characterized by the average values of the semantic features "Text emotionality" 0.7, "Bullying" 0.5, "Fear" 0.8, "Hate speech" 0.7 for the technique "Appeal to Fear-Prejudice" and "Text emotionality" 0.8, "Bullying" 0.4, "Fear" 0.7, "Hate speech" 0.6 are averaged-characteristic for technique "Exaggeration".

According to the values obtained by software, the results correlate with the values in table 1 with minor discrepancies. The discrepancies are also explained by the assessment of correlation of the test text to standard, since "Appeal to Fear-Prejudice" is estimated at 0.79, therefore the neural network has some doubts (in this example, the semantic feature "Text emotionality" is slightly overestimated at 0.9, compared to average value of 0.7) and "Exaggeration" with score of 0.68 (all semantic features differ from standard from 0.05 to 0.15). In general, proposed method demonstrates the high level of adaptability and efficiency, significantly increasing the accuracy of detecting most of the considered propaganda techniques and allowing for additional explanation of obtained neural network solutions. This makes it the promising tool for automated analysis of media content aimed at identifying and neutralizing political propaganda influences.

6. Conclusions

Method of semantic features estimation for political propaganda techniques detection using transformer neural networks was proposed, which allows to achieve an increase in explainability level of decisions made by transformer neural networks regarding the presence of political propaganda techniques by assessing semantic features manifestation, as well as to increase the accuracy of identification of gray and black propaganda techniques. This effect is achieved by using additional semantic features for political propaganda techniques and modified architecture of transformer neural networks that analyze not only text data, but also additional vector of numerical values of semantic features.

Applied study of developed method of semantic features estimation for political propaganda techniques detection using transformer neural networks revealed significant increase in the detection accuracy of 5 political propaganda techniques ("Appeal to Fear-Prejudice", "Repetition", "Causal Oversimplification", "Minimisation", "Appeal to Authority") and slight increase in the detection accuracy of 4 political propaganda techniques ("Red Herring", "Exaggeration", "Whataboutism" and "Thought Terminating Cliches"), the method was recognized as more effective for these techniques than existing analogues. For 8 political propaganda techniques ("Loaded Language", "Labeling", "Name Calling", "Black and White Fallacy", "Slogans", "Doubt", "Reductio ad Hitlerum", "Flag-Waving") detection accu-

racy increased slightly or did not increase, the method was recognized as parity for these techniques with existing analogues. For developed method, minimum accuracy of propaganda techniques detection is 0.85, maximum accuracy is 0.97, average accuracy is 0.89.

The developed method has significant potential for achieving the Sustainable Development Goals SDG No. 4 and SDG No. 16, in particular in increasing media literacy and critical thinking among the population. This will contribute to strengthening democratic institutions and ensuring transparency in political decision-making, which is important step in fight against disinformation and manipulation.

The directions of further research are to expand the set of detected semantic features to increase the accuracy of detecting political propaganda techniques, the accuracy of detection of which data by the method was determined to be insufficient. This demonstrates the potential for further optimization of the method for these specific political propaganda techniques.

Declaration on Generative AI: The authors have not employed any Generative AI tools.

References

[1] A. Horák, R. Sabol, O. Herman, V. Baisa, Recognition of propaganda techniques in newspaper texts: Fusion of content and style analysis, Expert Systems with Applications 251 (2024). doi:10.1016/j.eswa.2024.124085.

[2] G. Smyth, Nazi 'black' Propaganda to Britain: Secret Radio Stations and British Renegades, Historical Journal of Film, Radio and Television 44 (2024) 261–281. doi:10.1080/01439685.2023.2296215.

[3] S. Vajjala, Generative Artificial Intelligence and Applied Linguistics, JALT Journal 46 (2024) 55–76. doi:10.37546/JALTJJ46.1-3.

[4] A. A. Waoo, Sustainable development goals, Forever Shinings Publication, 2024.

[5] J. Szwoch, M. Staszkow, R. Rzepka, K. Araki, Limitations of Large Language Models in Propaganda Detection Task, Applied Sciences 14 (2024) 4330. doi:10.3390/app14104330.

[6] A. M. U. D. Khanday, Q. R. Khan, S. T. Rabani, M. A. Wani, M. ElAffendi, Propaganda Identification on Twitter Platform During COVID-19 Pandemic Using LSTM, in: A. A. Abd El-Latif, Y. Maleh, W. Mazurczyk, M. ElAffendi, M. I. Alkanhal (Eds.), Advances in Cybersecurity, Cybercrimes, and Smart Emerging Technologies, volume 4 of *Engineering Cyber-Physical Systems and Critical Infrastructures*, Springer International Publishing, Cham, 2023, pp. 303–314. doi:10.1007/978-3-031-21101-0_24.

[7] D. Chaudhari, A. V. Pawar, Empowering Propaganda Detection in Resource-Restraint Languages: A Transformer-Based Framework for Classifying Hindi News Articles, Big Data and Cognitive Computing 7 (2023) 175. doi:10.3390/bdcc7040175.

[8] M. Abdullah, O. Altiti, R. Obiedat, Detecting Propaganda Techniques in English News Articles using Pre-trained Transformers, in: 2022 13th International Conference on Information and Communication Systems (ICICS), 2022, pp. 301–308. doi:10.1109/ICICS55353.2022.9811117.

[9] D. G. Jones, Detecting Propaganda in News Articles Using Large Language Models, Engineering: Open Access 2 (2024) 1–12. URL: https://www.opastpublishers.com/peer-review/detecting-propaganda-in-news-articles-using-large-language-models-6952.html.

[10] G. Da San Martino, A. Barrón-Cedeño, H. Wachsmuth, R. Petrov, P. Nakov, SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles, in: A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, E. Shutova (Eds.), Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics, Barcelona (online), 2020, pp. 1377–1414. doi:10.18653/v1/2020.semeval-1.186.

[11] M. Abdullah, D. Abujaber, A. Al-Qarqaz, R. Abbott, M. Hadzikadic, Combating propaganda texts using transfer learning, IAES International Journal of Artificial Intelligence (IJ-AI) 12 (2023) 956. doi:10.11591/ijai.v12.i2.pp956-965.

[12] O. V. Barmak, O. V. Mazurets, I. V. Krak, A. I. Kulias, L. E. Azarova, K. Gromaszek, S. Smailova, Information technology for creation of semantic structure of educational materials, Proceedings

of SPIE - The International Society for Optical Engineering 11176 (2019) 1117623. doi:10.1117/12.2537064.

[13] G. Faye, B. Icard, M. Casanova, J. Chanson, F. Maine, F. Bancilhon, G. Gadek, G. Gravier, P. Égré, Exposing propaganda: an analysis of stylistic cues comparing human annotations and machine classification, in: V. Pyatkin, D. Fried, E. Stengel-Eskin, A. Liu, S. Pezzelle (Eds.), Proceedings of the Third Workshop on Understanding Implicit and Underspecified Language, Association for Computational Linguistics, Malta, 2024, pp. 62–72. URL: https://aclanthology.org/2024.unimplicit-1.6/.

[14] I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, E. Manziuk, O. Barmak, Method for neural network detecting propaganda techniques by markers with visual analytic, in: A. Pakstas, Y. P. Kondratenko, V. Vychuzhanin, H. Yin, N. Rudnichenko (Eds.), Proceedings of the 12th International Conference Information Control Systems & Technologies (ICST 2024), Odesa, Ukraine, September 23-25, 2024, volume 3790 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 158–170. URL: https://ceur-ws.org/Vol-3790/paper14.pdf.

[15] Y. V. Krak, O. V. Barmak, O. V. Mazurets, The practice investigation of the information technology efficiency for automated definition of terms in the semantic content of educational materials, Problems in programming (????) 237–245. doi:10.15407/pp2016.02-03.237.

[16] O. Zalutska, M. Molchanova, O. Sobko, O. Mazurets, O. Pasichnyk, O. Barmak, I. Krak, Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network, in: V. Lytvyn, A. Kowalska-Styczen, V. Vysotska (Eds.), Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems. Volume I: Machine Learning Workshop, Kharkiv, Ukraine, April 20-21, 2023, volume 3387 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 344–356. URL: https://ceur-ws.org/Vol-3387/paper26.pdf.

[17] I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, R. Bahrii, O. Sobko, O. Barmak, Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network, in: V. Vysotska, Y. Burov (Eds.), Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Systems. Volume III: Intelligent Systems Workshop, Lviv, Ukraine, April 12-13, 2024, volume 3688 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 16–28. URL: https://ceur-ws.org/Vol-3688/paper2.pdf.

[18] A. Paradise Vit, A. Magid, Differences in Fear and Negativity Levels Between Formal and Informal Health-Related Websites: Analysis of Sentiments and Emotions, Journal of Medical Internet Research 26 (2024) e55151. doi:10.2196/55151.

[19] H. Saleh, A. Alhothali, K. Moria, Detection of Hate Speech using BERT and Hate Speech Word Embedding with Deep Model, Applied Artificial Intelligence 37 (2023) 2166719. doi:10.1080/08839514.2023.2166719.

[20] D. Dimitrov, F. Alam, M. Hasanain, A. Hasnat, F. Silvestri, P. Nakov, G. Da San Martino, SemEval-2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes, in: A. K. Ojha, A. S. Doğruöz, H. Tayyar Madabushi, G. Da San Martino, S. Rosenthal, A. Rosá (Eds.), Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 2009–2026. doi:10.18653/v1/2024.semeval-1.275.

[21] P. N. Ahmad, Y. Liu, G. Ali, M. A. Wani, M. ElAffendi, Robust Benchmark for Propagandist Text Detection and Mining High-Quality Data, Mathematics 11 (2023) 2668. doi:10.3390/math11122668.

[22] J. Mitric, I. Radulovic, T. Popovic, Z. Scekic, S. Tinaj, AI and Computer Vision in Cultural Heritage Preservation, in: 2024 28th International Conference on Information Technology (IT), 2024, pp. 1–4. doi:10.1109/IT61232.2024.10475738.

[23] J. Choudhury, S. S. Gill, Performance Analysis of Machine Learning Models for Data Visualisation in SME: Google Cloud vs. AWS Cloud, in: S. S. Gill (Ed.), Applications of AI for Interdisciplinary Research, CRC Press, 2024, pp. 171–184. doi:10.1201/9781003467199-15.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, Édouard

Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

[25] A. Palomo-Alonso, V. G. Costa, L. M. Moreno-Saavedra, E. Lorente-Ramos, J. Pérez-Aracil, C. E. Pedreira, S. Salcedo-Sanz, TensorCRO: A TensorFlow-based implementation of a multi-method ensemble for optimization, Expert Systems 41 (2024) e13713. doi:10.1111/exsy.13713.

[26] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy, Nature 585 (2020) 357–362. doi:10.1038/s41586-020-2649-2.

[27] P. Gupta, A. Bagchi, Data Manipulation with Pandas, in: Essentials of Python for Artificial Intelligence and Machine Learning, Synthesis Lectures on Engineering, Science, and Technology, Springer, 2024, pp. 197–235. doi:10.1007/978-3-031-43725-0_6.