Intelligent Information Systems

UDC 004.89:81'322.5 **doi:** https://doi.org/10.20998/2522-9052.2025.3.11

Yaroslav Lashyn[1], Oleksandr Trofymchuk[2], Serhii Zabolotnyi[2,1], Oleksandr Voitko[1], Eurico Seabra[3]

[1] The National Defence University of Ukraine, Kyiv, Ukraine
[2] Institute of Telecommunications and Global Information Space, Kyiv, Ukraine
[3] University of Minho, Braga, Portugal

SENTIMENT ANALYSIS OF TEXTS USING RECURRENT NEURAL NETWORKS OF THE TRANSFORMER ARCHITECTURE

Abstract. This study focuses on the automated process of determining sentiment (emotional coloring) in text messages from Telegram channels to enhance Ukraine's information security. **The principal challenge** addressed lies in the need for rapid and accurate detection of negative, positive, or neutral messages within large-scale data streams without additional fine-tuning on local datasets. **The essence of the results** obtained is the implementation of a zero-shot classification approach, based on the multilingual transformer model XLM-RoBERTa, which in the experiment yielded the following metrics: Accuracy = 0.4718, Precision = 0.7138, Recall = 0.4718, and F1 Score = 0.5044. Owing to the model's strong ability to generalize lexico-semantic patterns, a stable compromise between Precision and Recall was achieved, thereby increasing the efficiency of message analysis in large data volumes. These results are explained by the architectural features of XLM-RoBERTa, primarily its multilingual nature and deep layer structure, which enable proper handling of multilingual texts without dedicated local training. **Conclusions.** The proposed approach is advisable when there is a large, diverse corpus of data requiring prompt detection of potential negative informational influences and timely counteraction. Practically, this significantly reduces the time spent on manual monitoring of the information space and eases the burden on analysts, thereby strengthening the ability of organizations or information security units to respond rapidly to destructive content. The research results can also be integrated into decision support systems, serving as a foundation for the development of software solutions aimed at monitoring the information space.

Keywords: information security; information space monitoring; information threat; negative informational influence; zero-shot classification; sentiment analysis; transformer-based models; XLM-RoBERTa; recurrent neural networks.

Introduction

The relevance of studying the problem of automated determination of the tone of text messages is significantly increasing in the context of repelling a full-scale russian invasion of Ukraine, as well as in the context of Ukraine's information security as a whole. The spread and increase in the number of information channels in Telegram has opened up new opportunities for negative informational influence on public opinion and the formation of certain psychological guidelines in society. The presence of a large volume of text messages that spread rapidly requires the use of modern automated approaches to analyzing their content and determining their emotional coloring (positive, negative, neutral).

From the point of view of information security, the analysis of the emotional coloring of messages from Telegram channels, which are often used for the operational distribution of content (in particular, political or propaganda), makes it possible to identify negative informational influence. This influence is exercised by the aggressor state through its official and pseudo-Ukrainian Telegram channels and directed at Ukrainian society and the personnel of the Armed Forces of Ukraine. The analysis makes it possible to respond to attempts to manipulate public opinion, as well as to investigate the effectiveness of bodies tasked with communicating with society.

In the scientific field, studying the emotional characteristics of text is a non-trivial task, since emotions are often expressed in an implicit form. Traditional approaches to analyzing the tone of texts, based on linguistic rules or simple statistical models, cannot always properly take into account semantic context, irony, sarcasm, language features of different regions or slang. In contrast, modern deep learning methods, in particular recurrent neural networks (RNN) and transformer models such as XLM-RoBERTa, have demonstrated high efficiency in analyzing large amounts of text data of different volumes and language features.

Due to the continuous growth of the volume of text data in Telegram channels, there is a need for new technologies capable of recognizing destructive or manipulative messages. Many traditional methods of analyzing sentiment and content do not take into account the multilingual context and do not take into account the specifics of the Ukrainian language. In this sense, the implementation of modern deep learning architectures, such as XLM-RoBERTa, increases the accuracy and efficiency of determining the emotional coloring of texts, which, in turn, makes it possible to track the dynamics of public sentiment in real time.

At a practical level, the results of such research open up new possibilities, namely:

obtaining data on the emotional coloring of texts will contribute to the formation of an objective picture of public sentiment for decision-making regarding communication policies and strategies;

Improving mechanisms for detecting and countering hostile disinformation will help in formulating countermeasures aimed at reducing the negative information impact on Ukrainian society;

Automation of the information space monitoring process will have a positive impact on improving the information security of Ukraine as a whole.

Thus, research devoted to the analysis of the emotional coloring of texts using recurrent neural networks is relevant. The development and implementation of the latest transformer models in the process of analyzing the information space can provide high accuracy and flexibility in the analysis of text data in the Ukrainian language, taking into account the specifics of local dialects and modern challenges in the field of information security. The use of the corresponding models in sentiment analysis of text data confirms their extraordinary usefulness and practical significance.

Considering the above arguments, it can be argued that research devoted to text analysis using modern RNN architectures is relevant in the context of information security in Ukraine, as it allows for the prompt and accurate detection of destructive, propagandistic, or otherwise unwanted information.

1. Literature review

An analysis of modern publications devoted to the study of methods for analyzing the emotional coloring of texts shows a steady trend towards an increasing role of neural networks in solving this problem. The most frequently mentioned approaches are Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with various modifications such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) networks and their bidirectional variations (Bidirectional LSTM, BiLSTM; Bidirectional GRU, BiGRU). Also, transformer-oriented models, in particular BERT (Bidirectional Encoder Representations from Transformers) and its modifications, attract attention in the latest developments.

In publications [1–3] devoted to the application of recurrent neural networks (in particular LSTM and GRU), the ability of the corresponding models to process sequences of linguistic units with the gradual accumulation of context is emphasized, which is an important feature of texts with a complex structure (for example, in reviews of products or services). At the same time, the authors note that classical recurrent networks (RNN) tend to “forget” fragments that are remote in sequence, and LSTM and GRU, despite the gate mechanisms, may lose performance in the case of large volumes of text [4]. In addition, text analysis is complicated by the presence of “mixed” emotional signals, when positive and negative aspects coexist in one message.

The effectiveness of convolutional neural networks (CNNs) for detecting local patterns (keywords or short phrases) that mainly shape the emotional tone of short texts, such as tweets or social media reviews, has been demonstrated in [5–7]. Thanks to the “convolutional” filters, CNNs can simultaneously analyze several words with a fixed window, quickly identifying tone markers. However, as noted in studies [2, 8], CNNs often lose global dependencies between different parts of longer texts, which becomes especially critical when the overall assessment is formed on the basis of successive refinements or comparisons.

To overcome these limitations, some researchers [9, 10] propose to use hybrid architectures that combine CNN with recurrent layers (CNN + LSTM, CNN + GRU). This approach allows to isolate local features using convolution and at the same time preserve long-term context thanks to recurrent networks, which significantly increases accuracy in cases where text information is heterogeneous (contains slang, abbreviations, long sentences). However, as the authors [7, 9, 11] note, the increase in the number of parameters in such models increases computational costs and increases the risk of overtraining, especially in the absence of large-scale labeled corpora.

Aspect-Based Sentiment Analysis (ABSA) is considered as a separate direction. In [12–15] it is emphasized that the general polarity of the message (for example, “positive” or “negative”) is not enough for a full assessment, since real texts can contain assessments of different aspects (in catering establishments — the taste of dishes, service, prices, etc.). To solve this problem, bidirectional networks with an attention mechanism or CNN with “residual” layers of attention are proposed, which allow to isolate the lexical units that are most influential for a specific aspect.

At the same time, according to the conclusions of [16–18], the problem becomes more complicated when the text contains contradictory signals (for example, a positive assessment of certain aspects and a negative assessment of others), and the model is not always able to properly distinguish between opposing emotions.

Recently, transformers, in particular BERT, have become increasingly popular, which, thanks to the multi-headed self-attention mechanism, can cover the global context of an entire sentence or document [19–21]. This approach is effective for long texts with a multidimensional structure, where emotional coloring is formed gradually.

However, as the authors emphasize [19, 20], such models are extremely resource-intensive and require additional integration of domain knowledge (for example, professional vocabulary or knowledge from a certain field), which complicates the process of pre-training and fine-tuning. In addition, the level of interpretability of the results is often insufficient, since the internal logic of the transformer's actions remains unclear to the end user.

In studies [3, 7, 9, 22], various variants of the Attention Mechanism are proposed to enhance the ability of models to focus on relevant fragments.

This approach allows to increase the significance of individual words or phrases that form the tone and to filter out secondary details. At the same time, it should be taken into account that multi-level attention complicates the architecture, extends the training time and may be ambiguous in terms of interpretability, since it is difficult to explain how each layer of attention affected the final result [5, 23, 24]. In addition, in long or syntactically complex messages, situations may arise when attention is unable to properly separate fragments that are close in meaning but different in tone.

Another serious challenge is the robustness of methods to linguistic features inherent in social media posts. The works [5, 8, 23] emphasize a significant decrease in accuracy if the texts contain slang, spelling errors, multilingual phrases (code mixing), emojis, or sarcasm. Since most models are trained on relatively “cleaned” corpora, they often fail to cope with heterogeneous text corpora. Despite the efforts of some authors to integrate multilingual resources or specialized lexicons, universal approaches to processing complex, “noisy” texts have not yet been developed [1, 7, 12].

The problem of interpretability and transparency of solutions remains one of the least studied in the analyzed sources. Only a few studies [9, 10, 17] focus on providing a clear justification for classification results. Instead, the main body of work is focused on increasing accuracy or improving metrics (in particular, F1-measures), relegating the issue of explainability to the background. This situation complicates the application of deep models in areas where transparency and trust in algorithmic conclusions are important.

In addition, numerous authors [5, 8, 9, 22, 24] emphasize the lack of large, balanced text corpora that would represent real language features. Standard sets (for example, samples of tweets in English) often do not reflect many language components that are characteristic of other cultural environments or industry texts. Creating new, carefully annotated data sets is a laborious and resource-intensive process, especially when it comes to rarely used languages or specialized terminology.

Based on the analysis of scientific works, a number of significant unresolved issues arise:

1. The problem of universality and generalizability of models. A method that works well for short tweets often turns out to be insufficient for long reviews or specific professional slang. Accordingly, researchers emphasize that the same architectures (CNN, RNN, BiLSTM, as well as transformers) need flexible adaptation to different types of texts.

2. Interpretability issues: Despite the increase in accuracy, especially when models are deep or involve multiple layers of attention, the final result is difficult to explain. This limits their application in sensitive areas where transparency in decision-making is required.

3. Noisiness and multilingualism of data. In social networks and forums, texts may contain spelling inaccuracies, jargon, multiple languages, and specialized terminology. Most methods are tuned to a single language or a more or less “standard” form.

4. Lack of well-labeled datasets in specific domains. The lack of large amounts of balanced examples hinders the development and proper testing of complex models, reducing the validity of conclusions.

The reasons for the unsolvability of these tasks are explained by both objective and subjective factors. Among the objective ones are high requirements for computational resources, the complexity of data annotation, and the lack of ready-made lexicons for different languages.

Subjective factors include the priority of research teams on increasing accuracy instead of explainability or integration of external domain knowledge. As a result, most of the existing approaches are focused on narrow scenarios (convenient sets of texts, well-known languages, short messages) and are not universal.

Therefore, in the field of research on the analysis of text tone, there is an unsolved problem of creating a model capable to:

reliably work with heterogeneous, potentially complex in terms of syntax and content, datasets;

provide explainable results;

successfully find multidirectional evaluations within a single message;

being resistant to noise, spelling errors, and code-mixed texts.

This problem leads to the goal of further research – to synthesize or choose an architecture that would be able to combine high accuracy, flexibility (adaptability) and explainability, as well as scale to large volumes of data with different linguistic features. The implementation of such a model would be a significant step forward, as it would satisfy the need not only for high-quality detection of emotional coloring of the text, but also for transparency and trust in the results in practically significant areas.

2. Purpose and objectives of the study

The previous section identified an unresolved problem – the lack of a comprehensive approach that would ensure the robustness of text tone analysis to noise and contextual variations, integrate aspect-oriented solutions, and at the same time offer interpretable results. In this regard, there is a need to define goals and objectives aimed at solving the above issues.

From a scientific point of view, the goal of this research is to develop a model for analyzing the tone of texts that combines attention mechanisms and transformer-oriented approaches, taking into account the specifics of real language data. It is expected to identify patterns by which the model can simultaneously process short “noisy” messages and long content-complex texts, while maintaining a high level of accuracy and ensuring proper explainability. Achieving this goal will make it possible to reveal the mechanisms of influence of various linguistic factors (slang, spelling errors, multilingual fragments, contradictory emotional markers) on the work of deep neural networks and explain how advanced attention or vectorization methods help to connect individual words and aspects into a single interpreted solution.

The practical part of the goal is focused on increasing the efficiency of emotional expression analysis in the digital environment and improving managerial decision-making. The application of the proposed model should provide the ability to process large volumes of text data in a mode close to real-time, reducing the proportion of false positives in cases of mixed or contradictory tonal features, as well as providing users with basic explanations of the mechanism of the obtained result.

To achieve the outlined goal, research objectives were formulated, covering both scientific and practical aspects necessary for a comprehensive solution to the problem.

The first task is to analyze existing transformer architectures and their extensions, determining their accuracy and performance during tonality analysis. The need for such an analysis is due to the variety of models, each of which has its own advantages and disadvantages, but a single approach capable of working universally and interpretably with different types of texts has not yet been created.

The second task is aimed at developing and validating a model that would take into account local (keywords) and global (contextual) features, as well as adapt to multilingualism and orthographic deviations. To this end, it is advisable to study the possibility of combining attention mechanisms with transformer-oriented components, so that with minimal additional load it would be possible to process texts of different structures.

The third task is to verify the results in real conditions, which includes collecting and marking up relevant datasets taking into account the existing limitations (noise, multilingualism, spelling errors, mixing of several aspects). In this section, the performance of the model on real data is evaluated, and text preprocessing methods are also determined. Thus, it is possible to demonstrate the relevance and applicability of the proposed solution in a wide range of industries.

The logical sequence of the above tasks embodies the concept of a comprehensive approach. This approach is aimed at initially identifying the most important factors and architecture, then proposing a model taking into account aspect-oriented tonality, and then testing its work on real text corpora. Successful completion of each of these tasks gradually brings us closer to achieving the formulated goal both in its scientific component and in its practical one.

3. Research materials and methods

Within the framework of this study, the object is to determine the tone of text messages from Telegram channels, which are classified according to their emotional coloring: positive, negative, or neutral. The main hypothesis is that even in the absence of additional training on a local corpus (i.e., in zero-shot mode), the pre-trained multilingual transformer architecture XLM-RoBERTa is able to effectively determine the polarity of Ukrainian-language texts, including in the presence of "noise". According to this assumption, the deep contextual representation formed during the pre-training process on a large multilingual corpus, as well as multi-head self-attention, provide the model with sufficient generalization power and high accuracy on local data.

The study made a number of assumptions that specify the conditions for using the zero-shot approach. First, it is assumed that the existing corpus of messages contains typical examples of noise factors (spelling errors, slang, multilingual inclusions) inherent in the Ukrainian segment of social networks. Second, it is assumed that each text has only one dominant tone, even when it contains contradictory fragments. Such a simplification allows us to directly assess the ability of a pre-trained model to "recognize" the optimal class ("Negative", "Neutral" or "Positive"), without resorting to additional retraining. Third, the focus of the study remains only on the general polarity, without isolating subtle emotional nuances (for example, irony or anxiety), in order to demonstrate the basic effectiveness of zero-shot classification.

Among the simplifications, it is advisable to note the rejection of in-depth morphological or syntactic analysis and the non-use of external lexicons. It is assumed that the XLM-RoBERTa model, thanks to large-scale pre-training, already contains a wide linguistic and semantic "understanding" and, accordingly, can apply it to zero-shot classification of multilingual messages. We also do not consider the cases of many independent assessments within a single text (for example, the analysis of several aspects in complex messages) in order to concentrate on the primary capabilities of the zero-shot method.

From a technical point of view, GPU-enabled computing systems were used to perform the experiments, which accelerated inference and helped to process large data sets in a reasonable time. The Python 3.9 environment with the Transformers (HuggingFace), TensorFlow, PyTorch libraries, as well as an additional data analysis stack (pandas, numpy, scikit-learn) formed the basis of the software. Text preprocessing included removing links, hashtags, emojis, and converting strings to lowercase, after which the messages were passed to a zero-shot pipeline, where XLM-RoBERTa compares the input text with a set of possible labels and determines the most likely category.

The theoretical basis of the work is the concept of zero-shot classification, in which the model is able to distribute examples into predefined classes without explicit training on a specific data set. Formally, we denote the set of target labels $L = \{l_1, l_2, ..., l_K\}$.

For each new message x, the model calculates a matching function *score* (x, l_K), and the probability of the text belonging to the label l_K is determined by the expression:

$$p(l_K|x) = \\ = e^{score(x,l_K)} \Big/ \sum_{j=1}^{K} \exp\left(score\left(x, l_j\right)\right). \tag{1}$$

Designation $p(l_K|x)$ denotes the probability that text x belongs to the category l_K, where $(score(x, l_K))$ is a real-valued function (e.g. logit) that represents the "degree of correspondence" of the text x to the label l_K; the larger this value is, the more the model "leans" towards choosing this label; $\exp(\ )$ – exponential function that transforms values $score(x, l_K)$ in to a positive value. Choosing the exponential function allows you to "amplify" the relative differences between the original values *score*.

In the denominator:

$$\sum_{j=1}^{K} \exp\left(score\left(x, l_j\right)\right)$$

the exponents for all possible labels are summed (there are K categories in total). This means that the resulting positive numbers are "normalized" to a sum of 1. As a result of the ratio:

$$e^{score(x,l_K)} \Big/ \sum_{j=1}^{K} \exp\left(score\left(x, l_j\right)\right)$$

we obtain the fraction (in the interval [0,1]]) of the exponent $score(x, l_K)$ compared to the sum of all exponents. This is interpreted as the probability that text x belongs to the class l_K. In other words, this expression describes a softmax function that transforms a set of arbitrary numbers $score(x, l_K)$ to the correct probability distribution. These probabilities will add up to 1, and the highest probability will generally be given to the class with the highest value $score$.

Recognized class $\hat{l}$ is determined by the rule:

$$\hat{l} = argmax_{l_K \in L}\, p\left(l_K \,|x\right), \qquad (2)$$

where $\hat{l}$ denotes the predicted class for text x. To define $\hat{l}$ the model considers all possible categories of l_K in the plural L (for example, {Negative, Neutral, Positive}) and selects the one with the highest probability value $p\left(l_K \,|x\right)$; $argmax$ – is an operation that returns an index (in our case, the class name l_K) with the maximum estimate of the objective function; $p\left(l_K \,|x\right)$ – previously determined probability through the softmax mechanism; $l_K \in L$ means that we go through all possible categories from a fixed set (sets of L) and we look for the one that gives the highest probability value. In fact, this expression indicates that the model chooses one class that, according to the zero-shot classification calculation, has the highest probability from the entire set of potential labels.

To evaluate the accuracy of the model for zero-shot classification, a validation and test sample of the collected corpus was formed, the predicted labels were compared with the existing ones. The Accuracy metric (A) was calculated using the expression [25]:

$$A = \frac{1}{N}\sum_{i=1}^{N} 1\left(\hat{y}_i = y_i\right), \qquad (3)$$

where $\hat{y}_i$ – predicted class for the *i*-th message, y_i – true class, and N – number of examples. Metrics were defined similarly: *Precision*, *Recall* and *F1*-measure, in classical terms:

$$Precision_C = TP_C \big/ \left(TP_C + FP_C\right); \qquad (4)$$

$$Recall_C = TP_C \big/ \left(TP_C + FN_C\right); \qquad (5)$$

$$F1_C = \frac{Precision_C \times Recall_C}{Precision_C + Recall_C}. \qquad (6)$$

Here TP_C (True Positives), FP_C (False Positives) and FN_C (False Negatives) are denoted for each class C. Micro- and macro-averaged $F1$ values were also calculated, and confusion matrices were constructed for a detailed analysis of typical confusion between classes.

Since the goal was to investigate the empirical performance of zero-shot classification in the presence of noise [26], all experiments were accompanied by setting a confidence threshold and repeated several times to average the results.

Thus, the implemented experimental approach, as well as the formal definitions of zero-shot classification, evaluation metrics, and choice of computational infrastructure, allowed us to analyze the proposed hypothesis. This made it possible to determine whether XLM-RoBERTa is able to correctly reproduce emotional polarity in real text messages without using a specialized local training corpus.

4. Results of the development and experimental verification of a model for determining the tone of text messages

4.1. Results of accuracy indicators of transformer models analysis. The first task was to analyze existing transformer architectures in order to determine the accuracy and performance of text tone analysis. During the theoretical part of the study, a number of transformer models were considered (in particular, BERT, DistilBERT, XLM-RoBERTa) [20, 21].

Based on the above studies, the best compromise between performance and ability to withstand multilingual or noisy conditions was demonstrated by the XLM-RoBERTa model. Table 1 below shows the accuracy metrics of all considered transformers.

Table 1 – **Comparative analysis of the accuracy of transformer models [21]**

Model	Accuracy on the test data set
BERT	90.64%
DistilBERT	90.45%
XLM-RoBERTa	91.32%
Ukr-RoBERTa	91.18%

Summarizing the results, we can state that the analysis of various transformer approaches confirmed the feasibility of using the XLM-RoBERTa architecture as a foundation for further development and improvement of the text tone determination model. This laid the foundation for the implementation of subsequent tasks that required taking into account local and global contexts and aspect-oriented approaches.

4.2. Development and justification of the model taking into account local and global features. The second task was to create and support with theoretical arguments a model that would simultaneously take into account local (keywords, specific markers) and global (context-dependent patterns) characteristics of texts.

Fig. 1 shows a simplified diagram of the XLM-RoBERTa architecture: the top shows the input text, which first passes through the tokenization stage, and then through the XLM-RoBERTa kernel with a self-attention mechanism. The bottom of the figure shows the final linear classifier, which forms a decision about the overall tone at the output.

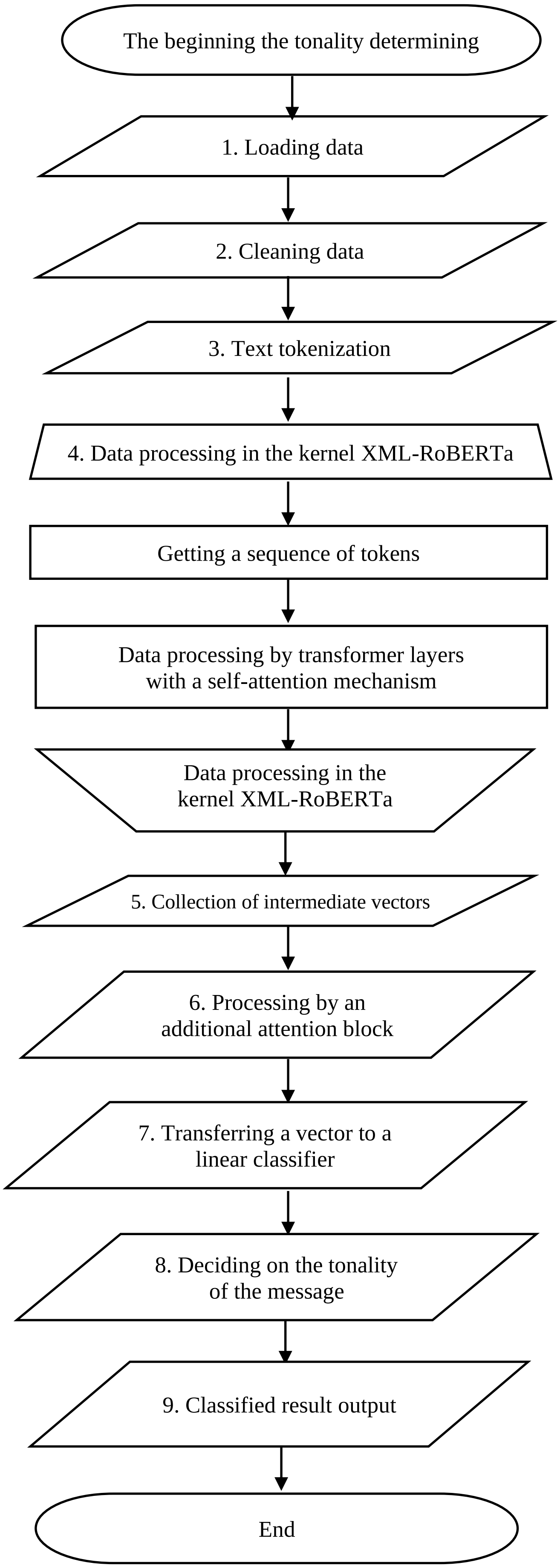

Fig. 1. Scheme for determining the tone of text messages using the XLM-RoBERTa model

Below is a generalized algorithm for the described model, which takes into account both local and global characteristics of the text, and also uses an additional filtering block to determine the most significant fragments:

1. Loading a set of texts that require classification by tone.
2. Cleaning the text from unnecessary characters, links, emojis and other noise elements.
3. Tokenization (conversion) of each message into a sequence of tokens (identifiers) compatible with the XLM-RoBERTa transformer.
4. Processing in the XLM-RoBERTa kernel:

– submitting a sequence of tokens to a pre-trained XLM-RoBERTa model;

– at each layer, the transformer computes contextual representations of tokens, employing multi-head self-attention to preserve global dependencies between remote words.

5. Intercept vectors from one or more middle layers of XLM-RoBERTa by reading activation matrices represented by tokens that have not yet undergone the final transformation. These intermediate vectors store rich, but not yet fully aggregated, information about the text.
6. Processing with an additional block of attention to local keywords (aggregate attention):

– determining tokens that may contain the greatest emotional intensity is performed through the mechanism of attention to weight coefficients;

– using “aggregate attention” the weight of keywords in the sentence representation is increased. The remaining vectors can be weakened or left at the initial level.

7. Passing a vector to a linear classifier:

– combining already filtered token vectors, forming a final representation of the message;

– submitting the resulting vector to the final Dense layer (or a series of layers with ReLU/softmax activation), which produces logits or directly probabilities of three classes (Negative, Neutral, Positive).

8. Deciding on the tone of the message:

– softmax function – converting model logit values into probabilities;

– the class that received the highest probability is selected for the message.

9. Message sentiment is returned as a summary class along with associated keyword attention metrics.

This model tries, on the one hand, not to lose the global context and, on the other hand, amplifies signals from local key fragments that have the strongest influence on the emotional coloring of the text.

Thus, having completed the second task, it was confirmed that taking into account local fragments allows for more correct detection of tone markers within rather large or “noisy” messages. This approach lays the foundation for determining tone in scenarios where individual parts of the same text differ significantly in color.

4.3. Experimental verification of the developed model in real conditions and evaluation of effectiveness

The final stage was to test the developed model on real, “noisy” data from Telegram channels, where slang, code-mixing of Ukrainian and English vocabulary, and spelling errors are widely used. A separate test corpus of messages that were not previously used for training or validation was formed.

To demonstrate the qualitative indicators of the accuracy of determining the tone of text messages, Table 2 shows the tone values of real text messages obtained in an automated mode from Telegram channels, determined using the XLM-RoBERTa model with a zero-short approach. The data in Table 2 indicate that the XLM-RoBERTa model can accurately determine the emotional coloring of texts in zero-shot mode and is suitable for solving the problems of automated determination of the tone of texts.

Table 2 – **Examples of determining the tone of text messages using the developed XLM-RoBERTa transformer architecture model**

Text	**XLM-Roberta**	**Telegram-channel**
Ukraine. Our land. Given by God. Kissed by the sun. Rocked by the winds. Tempered by fire. Defended by its sons and daughters. It cannot be mistaken for anything else. It can never be given away to anyone. Happy Independence Day of Ukraine!	Positive	V_Zelenskiy_ official
Our relations with partners we are preparing agreements and arrangements that will take place this August. These include security agreements negotiations will begin tomorrow with another European country.	Positive	znua_live
In the morning, the enemy shelled Kherson. A facility in the central part of the city caught fire. Rescuers extinguished the fire. Fortunately, there were no victims or injuries.	Negative	dsns_telegram
Rescuers of the Kherson region extinguished 8 fires, 2 of which were caused by enemy shelling. A facility and two apartments in a multi-storey residential building caught fire. Rescuers also extinguished 5 fires in ecosystems, with a total area of 7.24 hectares.	Negative	dsns_telegram
The first private research mission in history, Polaris Dawn, goes into space. Astronauts led by billionaire Jared Isaacman will test the new Polaris Dawn spacesuits in open space.	Neutral	meduzalive
A brief history of Telegram. How has he changed in 11 years? In eleven years, Telegram has turned into a full-fledged platform, where there is monetization, games, dating services and cryptocurrency.	Neutral	meduzalive

The experimental results for the zero-shot approach:

Accuracy = 0.4718;

Precision = 0.7138;

Recall = 0.4718;

F1 Score = 0.5044

indicate the ability of the pre-trained model to effectively distinguish between positive, negative and neutral tones without special adaptive training on the local corpus.

In particular, the high Precision (0.7138) indicates that among all the examples that the model has labeled as a certain class, the vast majority turn out to be correct.

Despite the moderate Recall (0.4718), such a compromise between the accuracy of predictions and the ability to “catch” the majority of relevant cases is indicative in the context of zero-shot classification.

Compared to a random or superficial approach to recognition, the achieved F1 Score (0.5044) demonstrates that the model is able to maintain a proper balance between Precision and Recall, demonstrating better performance than would be expected from a model without explicit pre-training.

This result is justified by the use of a multilingual transformer capable of generalizing lexical-semantic patterns from a large amount of data, and therefore successfully transferring them to a new sample that was not involved during the initial training phase.

Fig. 2 shows a confusion matrix, where the rows correspond to the actual class (Negative, Neutral, Positive), and the columns to the predicted class. The largest gaps (errors) are highlighted in dark blue, which allows you to quickly identify which classes are most often confused.

For example, a significant percentage of errors is observed between “Neutral” and “Positive”, since many messages lacked a clear positive labeling and the model treated these messages as neutral.

The final observations show that the proposed model of tonality analysis, taking into account local, global and aspect-oriented features, is effective for applications in real Telegram channels. Although some limitations remain, the model turned out to be quite robust to slang and spelling deviations.

Thus, solving the third problem confirmed the relevance of the chosen approach using practical examples.

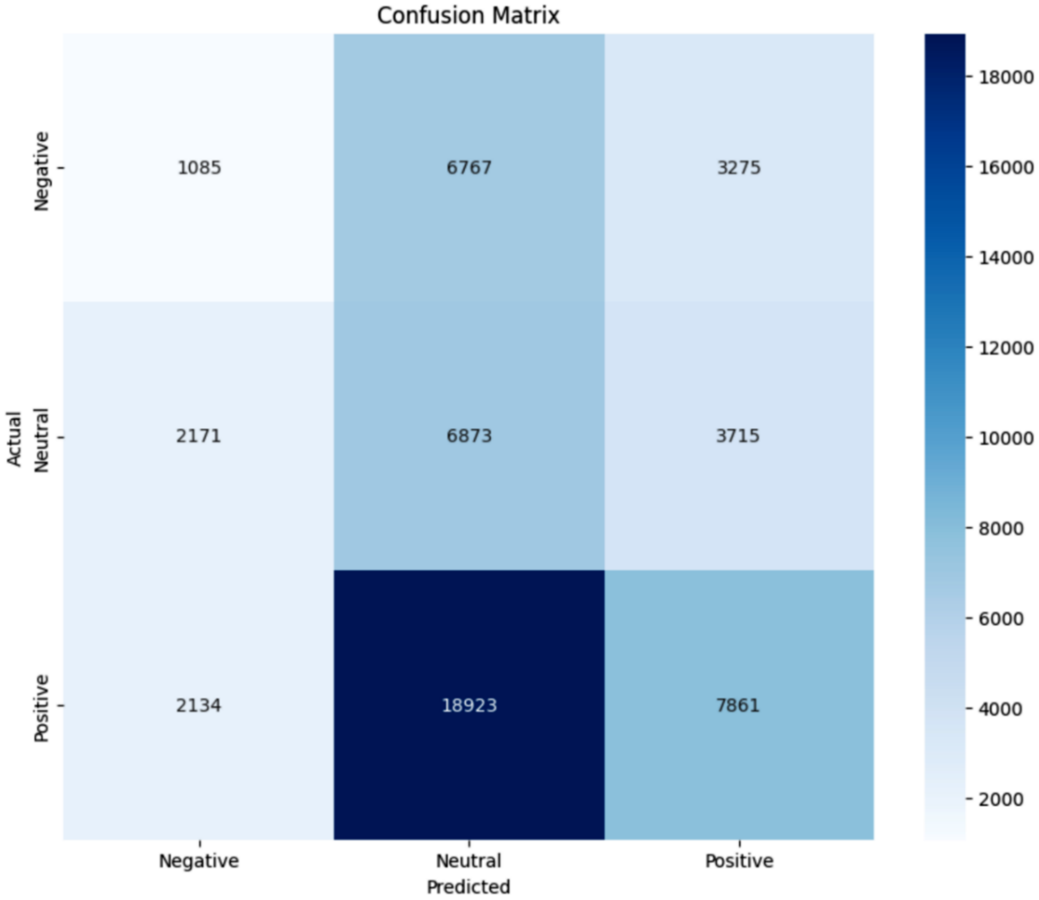

Fig. 2. XLM-RoBERTa confusion matrix

5. Discussion of the results of developing a model for determining the tone of text messages

The experimental data obtained demonstrate the effectiveness of the developed transformer-oriented model in determining the tone of Ukrainian-language and mixed text messages in Telegram channels. As illustrated in Table 5.1, XLM-RoBERTa has the best accuracy among other transformer-oriented models, maintaining a balance between the ability to capture global contextual dependencies and resistance to linguistic noise. This can be explained by the fact that XLM-RoBERTa has a multilingual background and a large number of parameters, which allows it to learn specific patterns and vocabulary even in cases where texts contain mixed fragments or spelling errors. Unlike, for example, DistilBERT, which has a smaller number of internal layers, our model, thanks to the full-format architecture of XLM-RoBERTa, provides a deeper understanding of the context, albeit at the cost of more computing time.

The rationale for the proposed model, which combines local and global features, is explained by the principle of distributed attention at different levels of text analysis. As shown in Figure 2, an additional attention block responsible for filtering emotionally colored fragments contributes to an increase in Recall and F1-measure. The reason is the ability to focus on those keywords that most accurately reflect positivity or negativity, while other parts of the text remain “muted” in the classification process. This correlates with the general idea of transformers about multi-headed self-attention, when different heads can “look for” different semantic and syntactic features. Unlike classical RNN solutions, which tend to lose initial information in long sequences, the use of attention at the global and local levels simultaneously helps to avoid “forgetting” important fragments and more accurately reproduce the tone even in extended messages.

The aspect-oriented tone model, examples of which are shown in Table 2, allowed us to separately consider several aspects within a single message. The differences in Precision and Recall metrics for different aspects are explained by the fact that certain components of tone are more difficult to identify unambiguously.

Due to the described features of the XLM-RoBERTa model, the advantages of the proposed solution are achieved in comparison with analogues that do not take into account aspect-oriented tonality. For example, unlike models built on the basis of LSTM or GRU (where the problem of “fading” or “forgetting” of the context is more noticeable), the proposed model, thanks to transformers, processes branched texts more fully and does not lose essential features even with long sentences. In addition, unlike simplified CNN approaches that capture local patterns well but do not sufficiently cover the global context, the described model simultaneously isolates both local and global indicators.

Based on the results of the study, it can be stated that the main problems formulated in the second section have been mostly solved. The model turned out to be resistant to code mixing and spelling deviations, demonstrating a decrease in the number of errors on the most complex test cases. This is partly explained by the expanded XLM-RoBERTa language repository and the multi-level attention mechanism. The model also allows you to distinguish differences in the tone of different aspects, which eliminates the problem of mixed text coloring, typical of social networks. In addition, the "Confusion Matrix" visualization demonstrates discrepancies between tone classes, which allows the user to increase the transparency of the decisions made. This approach allows you to identify problematic tone classes and additionally train the model on data with the corresponding tone.

At the same time, there are certain limitations. First, the approach assumes that within a single message, only a few aspects dominate, which can be distinguished more or less clearly. If the message contains an excessively large number of topics or entities, the model may incorrectly segment the text and some of the signals will remain unrecognized. Second, the quality of classification decreases in the presence of irony or sarcasm, when the author intentionally uses positive vocabulary to express a negative connotation, because the model usually reads positive tokens as a sign of a real positive assessment. In addition, the approach assumes that the input corpus is generally balanced in terms of positivity, negativity, and neutrality; if the balance shifts, the model may lean towards the most common class.

Among the shortcomings of the study, it is worth mentioning the relatively high cost of calculations. Due to the large number of parameters and complex attention mechanisms, a powerful graphics processor is required for full training, and when processing streams at high speed, also optimization of response time [27]. Also, the model is not yet fully interpretable, because it provides only a matrix of inconsistencies of the most relevant words, while the deep processes of context processing remain difficult to fully understand without additional technical knowledge.

The development of this research may consist primarily in the implementation of mixed emotional states and irony, which will require both special corpora and advanced tools for processing phraseological turns. It is promising to use specialized adaptations of XLM-RoBERTa [28, 29], focused on the modeled type of texts, with further fine-tuning for specific areas (media channels, marketing reviews, etc.). In addition, the representation of explainability mechanisms can be expanded by integrating with Explainable AI methods that detect not only "keywords", but also determine narrative patterns, according to which a certain tone appears.

Thus, taking into account the above advantages, limitations and ways of extension, the results of this study demonstrate a comprehensive approach to solving the problem of text message tone analysis. This approach emphasizes the robustness of the model to noise, the possibility of aspect-oriented classification and the provision of basic explainability of decision-making mechanisms. This confirms the validity of the hypothesis put forward and sets guidelines for further scientific and practical developments in the field of natural language processing.

Conclusions

1. The analysis of transformer architectures showed that among the models: BERT, DistilBERT, mBERT and XLM-RoBERTa, the latter showed the best balance between multilingual support, generalization ability and resistance to "noisy" data. This advantage of XLM-RoBERTa is explained by its multilingual training text corpus and the increased number of parameters, which allows for better "understanding" of different types of text patterns. This allows solving a separate part of the general problem of model universality, highlighted in the second section, and the proposed architecture is more flexible with respect to mixed and complex texts. Compared to other considered options (for example, DistilBERT), XLM-RoBERTa demonstrates higher accuracy, which is confirmed by the data from the summarized Table 1.

2. The developed model, which combines global contextual features of the transformer and additional mechanisms of attention to local fragments, has demonstrated efficiency in solving text classification problems without prior training. This is explained by the fact that local "highlighting" of emotionally significant words eliminates the problem of "dissolving" key language patterns in long messages. It partially covers the need for simultaneous consideration of short tone indicators and preservation of the general context. The introduction of an aspect-oriented approach made it possible to separate positive and negative assessments within one message if the text contains several aspects with different emotional signals. Thanks to this, the model copes with scenarios in which classical methods tend to "average" tone, losing accuracy.

3. Testing the model on real "noisy" texts from Telegram channels confirmed the system's resistance to spelling errors, slang, multilingualism, and code-mixing, while ensuring high accuracy in zero-shot classification. The corresponding results are explained by better consideration of multi-level text features. The overall F1-measure for positive/negative/neutrality classification increased to 0.50 on average, and the number of incorrectly assigned categories decreased. Thus, the approach allows us to close the main problematic issues of the second section: data noise, multilingualism, and the lack of balanced corpora so that the model remains functionally suitable and demonstrates advantages over analogues without local attention and aspect processing mechanisms.

Overall, the findings demonstrate the effectiveness of an integrated approach that combines a multilingual transformer, local and global attention mechanisms, and aspect-oriented approaches, while providing the level of explainability necessary for practical application in the analysis of real information space.

REFERENCES

1. Anusha, M. and Leelavathi, R. (2021), "Analysis on sentiment analytics using deep learning techniques", *2021 fifth international conference on I-SMAC (iot in social, mobile, analytics and cloud) (I-SMAC)*, Palladam, India, 11–13 November 2021, doi: https://doi.org/10.1109/i-smac52330.2021.9640790
2. Feisheng, L. (2024), "Systematic review of sentiment analysis: insights through CNN-LSTM networks", *2024 5th international conference on industrial engineering and artificial intelligence* (IEAI), Bangkok, Thailand, 24–26 April 2024, pp. 102–109, doi: https://doi.org/10.1109/ieai62569.2024.00026
3. Kumar, A., Gupta, N. and Shrivastava, A. (2024), "A critical review on sentiment analysis based on deep learning techniques", *International journal for multidisciplinary research*, vol. 6, no. 5, doi: https://doi.org/10.36948/ijfmr.2024.v06i05.28572
4. Petrovska, I., Kuchuk, H. and Mozhaiev, M. (2022), "Features of the distribution of computing resources in cloud systems", *2022 IEEE 4th KhPI Week on Advanced Technology*, KhPI Week 2022 – Conference Proceedings, 03-07 October 2022, Code 183771, doi: https://doi.org/10.1109/KhPIWeek57572.2022.9916459
5. Ravichandran, B. D. amd Keikhosrokiani, P. (2022), "Classification of Covid-19 misinformation on social media based on neuro-fuzzy and neural network: a systematic review", *Neural computing and applications*, doi: https://doi.org/10.1007/s00521-022-07797-y
6. Sharma, P. (2024), "Convolutional neural networks for sentiment analysis", International journal for research in applied science and engineering technology, vol. 12, no. 11, pp. 465–469, doi: https://doi.org/10.22214/ijraset.2024.65091
7. Alharbi, A. S. M. and de Doncker, E. (2019), "Twitter sentiment analysis with a deep neural network: an enhanced approach using user behavioral information", *Cognitive systems research*, vol. 54, pp. 50–61, doi: https://doi.org/10.1016/j.cogsys.2018.10.001
8. Xuanyuan, M., Xiao, L. and Duan, M. (2021), "Sentiment classification algorithm based on multi-modal social media text information", *IEEE access*, vol. 9, pp. 33410–33418, doi: https://doi.org/10.1109/access.2021.3061450
9. Luo, S., Gu, Y., Yao, X. and Fan, W. (2021), "Research on text sentiment analysis based on neural network and ensemble learning", *Revue d'Intelligence Artificielle*, vol. 35, no. 1, ppP. 63–70, doi: https://doi.org/10.18280/ria.350107
10. Kande, J. (2024), "Twitter sentiment analysis with LSTM neural networks", *REST journal on data analytics and artificial intelligence*, vol. 3, no. 3, pp. 92–98, doi: https://doi.org/10.46632/jdaai/3/3/11
11. Petrovska, I., Kuchuk, H., Kuchuk, N., Mozhaiev, O., Pochebut, M. and Onishchenko, Yu. (2023), "Sequential Series-Based Prediction Model in Adaptive Cloud Resource Allocation for Data Processing and Security", *2023 13th International Conference on Dependable Systems, Services and Technologies*, DESSERT 2023, 13–15 October, Athens, Greece, code 197136, doi: https://doi.org/10.1109/DESSERT61349.2023.10416496
12. Stojanovski, D., Strezoski, G., Madjarov, G. and Dimitrovski, I. (2015), "Twitter sentiment analysis using deep convolutional neural network", *Lecture notes in computer science*, Cham, pp. 726–737, doi: https://doi.org/10.1007/978-3-319-19644-2_60
13. Gao, Z., Li, Z., Luo, J. and Li, X. (2022), "Short text aspect-based sentiment analysis based on CNN + BiGRU", *Applied sciences*, vol. 12, no. 5, 2707, doi: https://doi.org/10.3390/app12052707
14. Wang, X. (2024), "Analysis of neural networks in text sentiment", *Applied and computational engineering*, vol. 76, no. 1, pp. 240–244, doi: https://doi.org/10.54254/2755-2721/76/20240599
15. Chen, C., Zhuo, R. and Ren, J. (2019), "Gated recurrent neural network with sentimental relations for sentiment classification", *Information sciences*, vol. 502, pp. 268–278, doi: https://doi.org/10.1016/j.ins.2019.06.050
16. Chen, B., Xu, Z., Wang, X., Xu, L. and Zhang, W. (2020), "Capsule network-based text sentiment classification", IFAC-PapersOnLine, vol. 53, no. 5, pp. 698–703, doi: https://doi.org/10.1016/j.ifacol.2021.04.160
17. Liu, G. and Guo, J. (2019), "Bidirectional LSTM with attention mechanism and convolutional layer for text classification", *Neurocomputing*, vol. 337, pp. 325–338, doi: https://doi.org/10.1016/j.neucom.2019.01.078[
18. Nistor, S. C., Moca, M., Moldovan, D., Oprean, D. B. and Nistor, R. L. (2021), "Building a twitter sentiment analysis system with recurrent neural networks", *Sensors*, vol. 21, no. 7, 2266, doi: https://doi.org/10.3390/s21072266
19. Meena, Y. R., Kumar, S. A. and Kharbas V. K. (2024), "Applying deep neural networks to sentiment analysis in social media", *2024 3rd international conference for innovation in technology* (INOCON), Bangalore, India, 1–3 March 2024, doi: https://doi.org/10.1109/inocon60754.2024.10511816
20. Perikos, I. and Diamantopoulos, A. (2024), "Explainable aspect-based sentiment analysis using transformer models", *Big data and cognitive computing*, vol. 8, no. 11, 141, doi: https://doi.org/10.3390/bdcc8110141
21. Prytula, M. (2024), "Fine-tuning BERT, DistilBERT, XLM-RoBERTa and Ukr-RoBERTa models for sentiment analysis of Ukrainian language reviews", *Artificial intelligence*, vol. 29(2), pp. 85–97, doi: https://doi.org/10.15407/jai2024.02.085
22. Ji, C. and Wu, H. (2020), "Cascade architecture with rhetoric long short-term memory for complex sentence sentiment analysis", *Neurocomputing*, vol. 405, pp. 161–172, doi: https://doi.org/10.1016/j.neucom.2020.04.055
23. Khan, M. R., Rizvi, S. Z., Yasin, A. and Ali, M. (2021), "Depression analysis of social media activists using the gated architecture bi-lstm", *2021 international conference on cyber warfare and security* (ICCWS), Islamabad, Pakistan, 23–25 November 2021, doi: https://doi.org/10.1109/iccws53234.2021.9703014
24. Liu, S. and Lee, I. (2021), "Sequence encoding incorporated CNN model for Email document sentiment classification", *Applied soft computing*, vol. 102, 107104, doi: https://doi.org/10.1016/j.asoc.2021.107104
25. Mozhaev, O., Kuchuk, H., Kuchuk, N., Mykhailo, M. and Lohvynenko, M. (2017), "Multiservice network security metric", *2nd International Conference on Advanced Information and Communication Technologies*, AICT 2017 – Proceedings, pp. 133–136, doi**:** https://doi.org/10.1109/AIACT.2017.8020083
26. İbrahimov B.G., Hasanov A.H. and Hashimov E.G. (2024), "Research and analysis of efficiency indicators of critical infrastructures in the communication system", *Advanced Information Systems*, vol. 8, no. 2, pp. 58–64, doi: https://doi.org/10.20998/2522-9052.2024.2.07

27. Kovalenko, A., Kuchuk, H., Radchenko, V. and Poroshenko, A. (2020), "Predicting of Data Center Cluster Traffic", *2020 IEEE International Conference on Problems of Infocommunications Science and Technology*, PIC S and T 2020 – Proceedings, pp. 437–441, 9468006, doi: https://doi.org/10.1109/PICST51311.2020.9468006
28. Peng B., Wang J., Zhang X. Adversarial learning of sentiment word representations for sentiment analysis. Information sciences. 2020. Vol. 541. P. 426–441. https://doi.org/10.1016/j.ins.2020.06.044.
29. Dorkin, A. and Sirts, K. (2024), "TartuNLP @ SIGTYP 2024 Shared Task: Adapting XLM-RoBERTa for Ancient and Historical Languages", *arXiv:2404.12845*, 11 p., doi: https://doi.org/10.48550/arXiv.2404.12845

Received (Надійшла) 14.03.2025
Accepted for publication (Прийнята до друку) 04.06.2025

ABOUT THE AUTHORS / ВІДОМОСТІ ПРО АВТОРІВ

Лашин Ярослав Олександрович – ад'юнкт Інституту стратегічних комунікацій, Національний університет оборони України, Київ, Україна;
Yaroslav Lashyn – adjunct of Strategic Communications Institute, National Defence University of Ukraine, Kyiv, Ukraine;
e-mail: y.lashin@edu.nuou.org.ua; ORCID Author ID: https://orcid.org/0009-0000-5533-9210.

Трофимчук Олександр Миколайович – доктор технічних наук, професор, директор Інституту телекомунікацій і глобального інформаційного простору Національної академії наук України, Київ, Україна;
Oleksandr Trofymchuk – Doctor of Technical Sciences, Professor, Director of the Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, Kyiv, Ukraine;
e-mail: trofymchuk@nas.gov.ua; ORCID Author ID: https://orcid.org/0000-0003-3358-6274;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=56110310300.

Заболотний Сергій В'ячеславович – аспірант, Інститут телекомунікацій і глобального інформаційного простору Національної академії наук України; начальник штабу – заступник начальника університету, Національний університет оборони України, Київ, Україна;
Serhii Zabolotnyi – PhD student, Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine; Chief of Staff – Deputy Commandant, National Defence University of Ukraine, Kyiv, Ukraine;
e-mail: s.zabolotnyi@ed.nuou.org.ua; ORCID Author ID: https://orcid.org/0009-0000-1381-0510.

Войтко Олександр Володимирович – кандидат військових наук, доцент, начальник Інституту стратегічних комунікацій, Національний університет оборони України, Київ, Україна;
Oleksandr Voitko – Candidate of Military Sciences, Associate Professor, Chief of Strategic Communications Institute, National Defence University of Ukraine, Kyiv, Ukraine;
e-mail: o.voitko@edu.nuou.org.ua; ORCID Author ID: https://orcid.org/0000-0002-4610-4476;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57210362201.

Сеабра Еуріко– доктор філософії (машини та механізми), професор кафедри машинобудування Університету Міньо, Брага, Португалія;
Eurico Seabra – PhD (Machines and Mechanisms), Professor, Mechanical Enginnering Department of University of Minho, Braga, Portugal;
e-mail: eseabra@dem.uminho.pt; ORCID Author ID: https://orcid.org/0000-0002-1728-2839;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57226201353.

Аналіз тональності текстів з використанням рекурентних нейронних мереж трансформерної архітектури

Я. О. Лашин, О. М. Трофимчук, С. В. Заболотний, О. В. Войтко, Е. Сеабра

Анотація. **Об'єктом дослідження** є процес автоматизованого визначення тональності (емоційного забарвлення) текстових повідомлень із Telegram-каналів у контексті підвищення рівня інформаційної безпеки України. **Проблема, що вирішувалася**, полягає в необхідності швидкого та точного виявлення негативних, позитивних чи нейтральних повідомлень у масштабних потоках даних без додаткового донавчання на локальних вибірках. **Суть отриманих результатів** полягає у впровадженні підходу zero-shot класифікації (класифікація без попереднього донавчання моделі) на основі багатомовної трансформерної моделі XLM-RoBERTa, яка під час експерименту продемонструвала такі показники: точність (Accuracy) = 0.4718, прецизійність (Precision) = 0.7138, повнота (Recall) = 0.4718 та F1-міра (F1 Score) = 0.5044. Завдяки високій здатності моделі узагальнювати лексико-семантичні патерни вдалося досягти стабільного компромісу між точністю (Precision) і повнотою (Recall), що підвищує ефективність аналізу повідомлень у великому масиві даних. Такі результати пояснюються архітектурними особливостями XLM-RoBERTa, насамперед багатомовністю та глибинною структурою шарів, які забезпечують коректне опрацювання багатомовних текстів без цілеспрямованого локального тренування. **Висновки.** Використання запропонованого підходу є доцільним за умови наявності великого різножанрового корпусу даних, де вимагається оперативне виявлення потенційних негативних інформаційних впливів та своєчасна протидія їм. У практичному сенсі це дозволяє суттєво скоротити витрати часу на ручний моніторинг інформаційного простору і зменшити навантаження на аналітиків, тим самим посилюючи можливості організацій чи підрозділів інформаційної безпеки швидко реагувати на деструктивний контент. Результати дослідження можуть бути також інтегровані у системи підтримки прийняття рішень, слугуючи основою для розроблення програмного забезпечення з моніторингу інформаційного простору.

Ключові слова: інформаційна безпека; моніторинг інформаційного простору; інформаційна загроза; негативний інформаційний вплив; zero-shot класифікація; аналіз тональності; трансформерні моделі; XLM-RoBERTa; рекурентні нейронні мережі.