FEDERAL STATE AUTONOMOUS EDUCATIONAL

INSTITUTION FOR HIGHER EDUCATION

NATIONAL RESEARCH UNIVERSITY

HIGHER SCHOOL OF ECONOMICS

Изображение выглядит как текст, Шрифт, логотип, Торговая марка

Автоматически созданное описание

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Diana Nazarova and Oleg Zubkov

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Master’s thesis**

**DEVELOPMENT OF MECHANISMS FOR IDENTIFYING CURRENT GEOPOLITICAL TRENDS THROUGH MEDIA ANALYSIS**

Field of study 01.04.02 Applied Mathematics and Informatics

Master’s programme “Applied Statistics with Network Analysis”

International laboratory for Applied Network Research

| Reviewer  Candidate of Sciences (PhD)  Galina Gradoselskaya  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Reviewer’s signature | Scientific Supervisor  Candidate of Sciences (PhD)  Daria Maltseva  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Supervisor’s signature |
| --- | --- |

Moscow, 2025

**The table of contents**

[**Annotation 3**](#_pb42p3y30kxk)

[**Introduction 3**](#_axescyty6s2b)

[**Theoretical framework 11**](#_2eije1lk7h5t)

[**Empirical part 18**](#_vb19vqx6y60b)

[**Conclusion 45**](#_gbd4vqzh1qqp)

[**Bibliography 49**](#)

# **Annotation**

In the 21st century known to be a digital era, the rapid development of the internet has transformed media into a vital and widely accessible source of information. For people around the world, it serves as the most useful and fastest way to stay informed and keep up with the global events. Yet, the media's role extends beyond simply delivering news; it also represents a vast repository of big data that can be systematically analyzed. It offers a great potential for identifying emerging geopolitical trends, it can be a great tool for forecasting. Such analysis of large volumes of information from the media can help to detect some unobvious tendencies and, in some cases, even help to predict the events - from presidential elections to international conflicts. This research aims to develop mechanisms for detecting relevant geopolitical trends through the application of advanced data analysis techniques, including text analysis, sentiment analysis and NLP.

Keywords: media, data analysis, geopolitics, geopolitical trends, machine learning, big data, text analysis, NLP.

# **Introduction**

***Background***

Everyday thousands of news articles are published on the internet. For people reading media is the easiest way to keep up with the political situation in their countries, global economic conjuncture and international relations in general. Media has a role of ecosystem where international conflicts, various views on political events, prospects of alliances and some changes in policies are covered and interpreted. Due to the endless amount of news, media data now serves as a rich and dynamic source which cannot be neglected in terms of analysis of the trajectories of geopolitical events (Benoit, 2020). At the same time there are some challenges: as the volumes of information in media are huge, and there is heterogeneity in data sources, highly sophisticated data analysis methods are needed as analytical tools.

For governmental structures, research institutions, policy makers and scientists in general the development of mechanisms with analytical approaches to identify some unexpectable tendencies, which will also help to generate some predictions in the geopolitical field, can also be highly advantageous and useful. In this work various data analysis methods will be applied to find out most appropriate mechanisms for media analysis aimed at geopolitical trends identification.

***Research problem:***

Although media analysis has been a relevant topic for many years and even decades, there is a lack of systematic tools and methodologies for identifying current geopolitical trends through media analysis. Existing studies often focus on isolated events, rely on limited data sources, or they don't have a universal algorithm for extracting topics and sentiment from article text. In this case, we solve this problem using open-source libraries and the GNews API. Some of the sources found during the process of literature review, at first glance seemed to be relevant as they conducted real-time analysis, however, they stopped being updated several years ago (such as The GDELT Project)[[1]](#footnote-0). This research aims to fill this gap by developing a prototype mechanism that utilizes machine learning methods, especially sentiment analysis and thematic classification via NLP, to process media content and identify relevant geopolitical patterns in selected case studies such as BRICS and Russia. It should also be mentioned that the existing methods of political trends detection based on media analysis are mostly represented in foreign sources while in the Russian scientific base it can be a challenge to find high quality research related to this topic. The result of this scientific paper will be represented as a guide consisting of all the stages needed to conduct analysis of mass media aimed at geopolitical tendencies detection. Especially at this time of political instability, our master’s thesis has a great potential to be practically-oriented and valuable for research institutions and even governmental structures as a base for the future studies or for the elaboration of a real-time base aimed at processing the topic request and providing up-to-date results analyzing available news sources.

***Object of research:***

The object of research is modern geopolitical trends as reflected in media discourse on specific key topics.

***Subject of research:***

The subject of the research is practical methods for detecting and analyzing geopolitical trends through automated media analysis using NLP and sentiment scoring.

***Empirical object:***

Database of news articles related to politics, international relations, etc. (title, description, content, URL, the source name where the news article is published, source URL).

***Data collection:***

Database for this research was gained through GNews API – it allows the download and processing of news articles from more than 60 thousand sources.

***Research tasks:***

1. Conduct a literature review on existing approaches to media-based geopolitical analysis, with a focus on the real cases where data analysis methods helped to obtain some crucial results.
2. Formulate the cases which will serve as examples for testing the algorithm.
3. Collect and preprocess the scope of news articles with the use of GNews API for the several cases which will be considered in this research.
4. Apply natural language processing (NLP) techniques to identify the key topics and sentiment trajectories within the media texts.
5. Compare media narratives across different countries.
6. Assess the strengths and limitations of the applied analytical methods.
7. Evaluate the predictive potential of the developed mechanism.
8. Formulate the algorithm with step-by-step actions needed to conduct media analysis and provide conclusions.

***Limitations of the study***

This study is subject to several limitations:

* Language bias: although multilingual models are applied, the majority of high-quality NLP tools are adapted to the English language, which is likely to have a negative impact on performance on non-English sources – in some cases it is needed to manually translate news articles before conducting automatic analysis.
* Data source disparities: media platforms differ in editorial policies, tone, and audience engagement strategies. The generalizability of results may be affected by differences in media system structures.
* Event context sensitivity: some geopolitical narratives are deeply embedded in local historical and cultural contexts, which is hard to take into consideration when conducting just computational analysis.
* Black-box models: while many ML and NLP tools function as black boxes, our research uses very specific and open models, such as TextBlob, and also uses thematic classification by keywords, which ensures the transparency and reproducibility of the created algorithm.

***Ethical Concerns***

All media data used in this study is publicly accessible and obtained through legal and transparent means as these news articles are published on the internet and are free to use. The analysis is conducted in compliance with ethical research standards, including:

* No personal data was collected for the research: the dataset consists solely of published news content and does not include user comments or private information.
* Source transparency: all media sources are documented, and no misrepresentation or manipulation of texts is performed.
* Bias awareness: analytical results will be interpreted with attention to potential ideological and cultural bias in the media sources themselves, the results will be formed as neutral description of the product.

**Methodology of research**

The research methodology is built as a pipeline consisting of several interconnected steps combining qualitative and quantitative approaches. The steps that make up the pipeline are:

1. Collecting information using the GNews API.
2. Extracting the full text of articles using the newspaper3k library.
3. Preprocessing: text cleaning, segmentation, normalization.
4. Thematic analysis with dividing sentences into bigrams and calculating the frequency of occurrence of bigrams with their correlation to certain thematic groups.
5. Sentiment analysis using the TextBlob library
6. Aggregation of results for each country and visualization.

One of the main focuses of this research will be made on *sentiment analysis* conducted through machine learning. Sentiment analysis is a crucial tool for detecting the general direction of changes in countries, the views of populations, authorities or scientists on global events, on political reforms, etc. The use of sentiment analysis is represented in research by Wang, Wang, 2023 giving insights into positive changes in China caused by the softening policy of the Chinese government in terms of Covid – 19. In the article by Ortiz, Rodrigo, Sicilia, (2023) sentiment analysis helps to indicate the status of international conflicts according to “signals” in media by country.

Machine learning is actively applied in research related to politics. One of the examples can be an article by Ortega‑Martorell, S., Bellfield, R. A. A., Harrison, S., Dyke, D., Williams, N., Olier, (2023), which uses machine learning methods to conduct clusterization of countries in accordance with the level of freedom of speech. Natural language processing as a machine learning algorithm is a powerful instrument for analyzing big text data (Albanese et al., 2020). For example, natural language processing (NLP) techniques are used to analyze data related to the US withdrawal from Afghanistan in 2021. The study shows that NLP analysis can detect polarized opinions, measure emotional reactions to geopolitical events, and reveal prevailing narratives (such as support for Afghan civilians or criticism of US policy). The method improves the granularity of trend detection by combining geographic and temporal metadata, allowing for real-time tracking of reactions around the world (Anwar et al., 2022).

# **Theoretical framework**

The interplay between media and geopolitics represents a dynamic and compelling field of study for scholars from a variety of disciplines. Policymakers, corporations, journalists, and the general public are all connected by the media, which plays a vital role in promoting information sharing and influencing public opinion. The media has many functions in contemporary society, including reporting on political events and economic reforms, influencing geopolitical trends, and even influencing voter behavior (Strömberg, 2015).

Media coverage is crucial in shaping political narratives, frequently influencing how audiences prioritize and comprehend issues, according to research by University of Washington, (2023). Concentrating on specific international disputes or economic policies covered in the news, for example, can direct public and political opinion onto specific international matters. Online media channels have revolutionized the way information diffuses at high speeds and across large geographies. The study by Boydstun, (2022) highlights that online news sources have turned out to be powerful political leaders' tools to reach the masses directly. In the book by Brennen, (2022), the media is framed as a product of specific political conditions that “offers insights into societal context”.

Nevertheless, this research will focus on the primary role of media as a source of information and on the ways to use it for analysis and detection of geopolitical tendencies.

**Review of literature dedicated to general topic**

Quite a big number of articles are dedicated to the ways media analysis helps to determine most important up-to-date tendencies in the field of geopolitics and even to make some relevant predictions. In this part of the master thesis various example of such analysis will be taken into account, and attention will be paid to the methods of data analysis applied in these pieces of research. For instance, one of the main methods of data analysis – machine learning – is actively used during the last decades in the field of political sciences, particularly in conflicts study (de Slegte, Van Droogenbroeck, Spruyt, Verboven, & Ginis, 2024).

The development of mechanisms for identifying geopolitical tendencies can be a useful tool in different areas: conflicts and crises prediction, identification of a country’s increasing leadership, forming alliances in geopolitics or detection of the changing moods or opinions about some political events, countries’ policies or even in relation to some political leaders. Big data use in these fields is not very new, however, there is always a space for progress as technology and Artificial Intelligence is developing dramatically fast, so there are more ways to apply quantitative data analysis. The advancement of Large Language Models (LLMs), in particular, has the potential to transform research methodologies and the way big data analysis is used (Ortiz, Rodrigo, Sicilia, 2023). One of the representative examples is Integrated Crisis Early Warning System (ICEWS) used by the organization CrisisWatch – the tool is aimed at identifying the tendencies of current political conflicts and with the use of different criteria the system is able to determine rising tensions between countries, escalation of internal conflicts – generally it tracks the situations in a great number of countries which may allow to prevent some most dangerous courses of events (Arradon, 2024). The most important issue of combining big data analysis and media is the opportunity to take early actions when analytical algorithms determine the signs of exaggeration of the conflicts. At the same time the author of the article highlights that despite all the existing opportunities of deep data analysis, a human-led approach is needed to make proper conclusions (Arradon, 2024). Another study focusing on preventing conflicts is an article related to identifying geopolitical event precursors (Ramakrishnan et al., 2022). The method of attention-based long short-term memory (LSTM) uses text datasets such as news articles and social media with the detection of word context patterns to elaborate predictions on political conflicts. As examples the study uses the protests in Latin America and the Middle east conflict.

The article “Media in the Geopolitical Crossfire” (Puhr & Kupfer, 2023) describes the idea of media analysis as a useful tool to study policies conducted in different countries, however suggesting that media tends to be biased and proposing not only paying attention to new articles, but also considering some neutral sources as Google trends[[2]](#footnote-1), for example. Another study (Lin, Chichen, Yongbin Wang, Chenxin Li, Weijian Fan, Junhui Xu, and Qi Wang, 2024) considers media profiling as a critically important tool for understanding global information landscape, however, as news articles are published in different languages, the article proposes that the only way to analyze media comprehensively is to apply automated analysis.

The paper "A New Generation of Conflict Prediction" offers a new approach to predicting political conflict by combining machine learning and the analysis of event data. The authors note that traditional conflict prediction models have generally relied on structural variables (e.g., demographic or economic variables), and they have therefore failed to capture dynamic, short-term changes in levels of risk of conflict. The authors offer a new framework which leverages high-frequency event data harvested from the media, which they process via advanced computational techniques to produce predictions of conflict with finer temporal, spatial and predictive accuracy. This paper proves that civil conflict can be predicted with greater predictive accuracy in regard to its onset and escalation and supports the argument by highlighting the usefulness of event data using the media to observe the dynamics of geopolitical trends in real time (Kaya, 2024).

The study conducted by the OECD focuses on the idea of analyzing social media in order to uncover the most relevant topics raised by people – users of social media, key tendencies and the ways to use these results for implementation of reforms, some developments and social changes in government. The article suggests that the governments could make improvements on the basis of these results. In their research Strand and Fjelde (2019) consider text as data: the news articles in combination with texts of speeches are subject to analysis aimed at identification of the mail political preferences of the population, leading ideologies and dynamics of changes in geopolitics.

Research conducted by Dubois and Blank (2018) emphasizes the necessity for data-driven methodologies that utilize trace data to improve the reliability of trend identification systems. It is highlighted that for obtaining relevant results the news sources must be diverse and independent from each other (Graff et al., 2025), which will also serve as principle for the empirical part of this master’s thesis.

**Text analysis and forecasting**

Many articles are dedicated to forecasting based on different types of text analysis: The study by Burns, Kelsey & Donovan (2025) provides an overview of the real-time analysis of posts in social media and was aimed at creating real-time “geopolitical topic modelling”. According to the results of the authors, the model was able to predict most relevant topics faster than Google Trends[[3]](#footnote-2) algorithm. Scientific paper published by Mueller and Rauh (2013) also focuses on predicting geopolitical tendencies, in particular, political and social conflicts applying machine learning analysis to the news articles published by the media. It is highlighted that generally there is a global necessity to develop an “early-warning system”. In the study not only the text of news articles are used as data for analysis, but also economic indicators are taken into consideration which creates a comprehensive approach to the problem. However, the authors highlight that there are some limitations such as the difficulty to generate holistic dataset and model.

**Cases with media-based analysis aimed at identifying political tendencies in different countries**

It is highly crucial to consider here several real cases where media analysis helped to detect some political tendencies in different countries. In the article by Hinck, R from Texas A&M University Russian media is reviewed in order to get an understanding on whether Russia and the USA have some common interests. According to the author, after the analysis of more than 2,5 thousand news articles, Russia represents itself in the media as the state which can oppose the corrupted western world thanks to such organizations as BRICS or EEU (Eurasian Economic Union). The results show that Russia actively covers the topic of NATO in the media, resorting to inconsistencies between NATO countries thereby putting the U.S. at disadvantage. (Hinck, R. et al., 2016).

The paper titled "Measuring Geopolitical Risk" authored by Dario Caldara and Matteo Iacoviello, which appeared in the American Economic Review, presents a news-based Geopolitical Risk (GPR) index designed to measure the adverse effects of geopolitical events and their economic repercussions. This index is derived from automated text analysis of news articles and exhibits significant increases during major geopolitical occurrences such as the World Wars, the Korean War, the Cuban Missile Crisis, and the aftermath of 9/11. The research indicates that elevated geopolitical risk is associated with a decline in investment and employment, a heightened likelihood of disasters, and increased economic downside risks, influenced by both the potential and actual occurrence of geopolitical events. Furthermore, the authors create GPR indicators at the industry and firm levels, demonstrating that sectors and companies facing greater geopolitical risk encounter more substantial reductions in investment (Caldara, Matteo, 2022).

The example of use of the sentiment analysis for identifying trends in media is represented in the research by Imran et al. (2020). The authors came to conclusion that Sweden considered Covid-19 in much more positive way in comparison with Norway, moreover, that there were “strong sentiment correlations” between the USA and Canada – although this topic is not about geopolitics, it gives an understanding that sentiment analysis can easily be applied to various cases which need to be explored.

To sum up everything that has been stated so far in this part of the master’s thesis, the literature review demonstrates the increasing academic and practical interest in the utilization of media content as a basis for understanding and predicting geopolitical change. As the media platforms have been reconstituted into rich ecosystems of information transmission, the platforms are now performing not only as conduits of communication but also as mirror-like reflections of global political change, ideological alignments, and territorial conflicts. In this kind of approach, the examination of media texts offers a significant methodological direction for stressing the subtleties encompassed in geopolitical discourses and evaluating public as well as institutional reactions to new events.

Among the main findings of the literature review is the identification of media as a source of real-time information and as an instrument of influence in international relations.

It plays a dual role: first, by documenting geopolitical events in real-time; and second, by shaping perceptions, narratives, and policy debates around those events. This twofold function justifies the place of media analysis within the broad scope of geopolitical research. As the media landscape continues to be digitized and decentralized, the potential for gleaning intelligence from this vast repository of text has grown significantly. The other key theme that is present is the usefulness of applying computational techniques—i.e., natural language processing, sentiment analysis, and topic modeling—to big datasets of media texts. These technologies allow for the efficient processing of huge volumes of information, permitting researchers to reveal concealed patterns, new topics emerging, and sentiment changes along time spans. Compared to conventional manual content analysis techniques, computational techniques provide greater scalability, objectivity, and reproducibility and are thus especially well-suited to the high-volume and ever-changing nature of modern media.

In spite of numerous benefits, existing literature highlights several methodological issues. One issue concerns linguistic constraint with various analysis tools, which are frequently tailored to one prevailing language, thereby excluding non-English text and introducing possible biases.

A different problem is the explainability of sophisticated machine learning models, which, although they do possess excellent capabilities, sometimes operate as "black boxes" so that it is hard to discern how conclusions are reached depending on the input data. In addition, geopolitical discourses tend to be founded on culturally relative metaphors, historical backgrounds, and rhetorical techniques that are not necessarily caught by purely algorithmic approaches. There is, nonetheless, a firm academic consensus on the potential of media-based analysis to enable trend identification, monitor conflicts, and establish early-warning systems. The field is also fast developing, with technological progress bringing more complex solutions to existing limitations. Incorporation of multilingual models, contextual embeddings, and hybrid approaches blending qualitative and quantitative aspects are poised to significantly expand the effectiveness and feasibility of media analysis in geopolitical research. In summary, the studied literature provides a strong justification for media content being used as the foundation of analyzing geopolitical trends. Furthermore, it highlights the necessity of creating systematic, unambiguous, and flexible methodologies that are capable of functioning effectively across various linguistic and cultural settings. The findings provide a conceptual and theoretical basis for the empirical component of this research, in which an integrated system for detecting geopolitical trends using computerized media analysis is developed and tested.

# **Empirical part**

The goal of the project was to develop a mechanism for analyzing media content from news sources in order to identify current geopolitical trends emerging in the international agenda. The main focus of the analysis was to study the subject matter (dividing each news item into the topics it covers) and the tone of publications (i.e. how exactly the publication is written - in a positive, negative or neutral tone).

During the development and testing of the mechanism, two keywords were used to search for news:

1. “brics”, to filter all news related to this organization;
2. “Russia”, to filter news related to Russia.

**BRICS**

The general algorithm consists of six main stages:

1. Data collection using GNews API.
2. Obtaining full texts of articles via URLs available after collection.
3. Text cleaning and thematic markup of articles.
4. Thematic classification and sentiment analysis.
5. Aggregation and analysis of the obtained indicators by country.
6. Visualization of the obtained data.

We will tell you more about each stage.

**Data collection**

The goal of this stage was to collect news articles using the GNews API by the keyword “brics”. The choice of this keyword (short for Brazil, Russia, India, China, South Africa) is a conscious and strategically sound step to develop a mechanism for analyzing current geopolitical trends. BRICS is an alternative to the Western-centric model of global governance. This organization has become one of the key links in the global politics of the multipolar architecture of world governance, which forms the modern balance of power between countries and regions. In 2023, the BRICS expansion process began. Iran, Egypt, Ethiopia, Saudi Arabia\* and Argentina became the members of BRICS), which caused a serious resonance in the world media and made this association a marker of global changes.

In addition, this word is universal for news in different languages, which is a useful factor for news selection. This word is also often mentioned in news articles:

* about geopolitical alliances;
* about the global economy;
* about energy and finance;
* about initiatives in the field of technology, trade, security, etc.

It is for these reasons that BRICS was chosen as the first entry point for developing the mechanism.

So, the first stage is data collection.

We used the GNews API and the capabilities of this tool to obtain basic information about news by certain keywords with certain parameters. As parameters, only the language parameter was changed - news was selected exclusively in English.

In addition, GNews only supports a strictly defined list of countries, the list of which with a two-letter ISO code is presented in the table below:

| Name | ISO Value |
| --- | --- |
| Australia | au |
| Brazil | br |
| Canada | ca |
| China | cn |
| Egypt | eg |
| France | fr |
| Germany | de |
| Greece | gr |
| Hong Kong | hk |
| India | in |
| Ireland | ie |
| Israel | il |
| Italy | it |
| Japan | jp |
| Netherlands | nl |
| Norway | no |
| Pakistan | pk |
| Peru | pe |
| Philippines | ph |
| Portugal | pt |
| Romania | ro |
| Russian Federation | ru |
| Singapore | sg |
| Spain | es |
| Sweden | se |
| Switzerland | ch |
| Taiwan | tw |
| Ukraine | ua |
| United Kingdom | gb |
| United States | us |

The basic query downloads news for each of the countries represented, taking into account that they must be in English. As soon as all the news is received in the temporary dictionary list, the file “XX\_articles.json”, where XX is the two-letter ISO code of the corresponding country. The uploaded news data has the following format (example):

{

"*title*": "Nigeria’s Brics partnership: economist outlines potential benefits",

"*description*": "Nigeria can benefit from a Brics partnership but it must find a careful balance between the relationship and the interests of its western allies.",

"*content*": "During its 16th annual summit in Kazan, Russia, Brics – a group of emerging economies determined to act as a counterweight to the west and to whittle down the influence of global institutions – invited Nigeria and eight other countries to join it as ... [6692 chars]",

"*url*": "https://theconversation.com/nigerias-brics-partnership-economist-outlines-potential-benefits-248943",

"*image*":"https://images.theconversation.com/files/646876/original/file-20250204-16-k2a140.jpg?ixlib=rb-4.1.0&q=45&auto=format&w=1356&h=668&fit=crop",

"*publishedAt*": "2025-02-04T13:00:00Z",

"*source*": {

"*name*": "The Conversation UK",

"*url*": "https://theconversation.com"

}

As it can be seen, we get quite comprehensive information on each news item – title, short description, first N characters of the news item itself with the total number of characters in it, URL where the news article is located, image (if any), Date and time of publication in standard ISO 8601 format always in UTC time zone, as well as brief information about the source – source name and source home page.

Since not all countries have predominantly English-language news, we got a list of the following countries:

1. Australia
2. Canada
3. Switzerland
4. United Kingdom
5. Ireland
6. India
7. Philippines
8. Pakistan
9. Singapore
10. USA

So, as you can see, there are 10 countries. Let's move on to the next point.

**Get full text of news articles**

At this stage, a file was prepared that gets the full text of the news using the functionality of the newspaper3k library. The use of this library is justified given its functionality, efficiency and reliability in the context of parsing web pages of news resources.

The problem was that, as you may have noticed, GNews only gives a truncated text of the news. This makes it impossible to:

* classify by topic (it is impossible to highlight all the topics covered in the truncated news);
* analyze rhetoric (the truncated part may contain important semantic structures for analysis);
* identify trends in content (the truncated version of the news will not allow you to analyze its overall tone, since most often the news begins with an introduction and a brief overview of what will be discussed in it).

And therefore, since our task is to analyze the entire news, we needed to get the full text of each news. But manually copying a large volume of news is counter-effective, and that is why the capabilities of the newspaper3k library were used.

This library allowed:

1. Download an HTML page;
2. Recognize the main article (without sidebars, ads, etc.);
3. Save the resulting article under the key “full\_content” in the dictionary of each news item in each country.

Let's move on to the next step.

**Cleaning and thematic markup**

The next stage is cleaning and thematic markup. At this stage, the key task of moving from the "raw" text of news articles to a structured array suitable for further thematic analysis, as well as determining the general tonality of each news item, is implemented. The goal of this stage was to:

1. Clean the text from HTML tags, links, symbols unnecessary for analysis;
2. Break the text into separate sentences.
3. Highlight key topics by bigram frequency.

The first stage is cleaning. For this, the BeautifulSoup library was used, which can recognize HTML and XML tags well and perform certain manipulations with them depending on the task. Our task is to remove these tags, since they are not needed for analysis.

Next, the text is cleaned using regular expressions from unnecessary characters, spaces (double, if any) and links like "https:".

After that, the text is broken into separate sentences (segmented) using nltk.sent\_tokenize. This is done to enable a more subtle analysis of the sentiment, i.e. by phrases or small blocks (sentences), rather than by the entire text at once, and also allows for a higher accuracy of thematic division of the entire news item into thematic tags that prevail in it.

The next stage is to extract topics using bigrams, i.e. pairs of words that can be interrelated. For example, the phrase “mama washed the frame” contains two bigrams – “mama washed the frame” and “washed the frame”. Using frequency bigrams will allow us to better classify the text by topic. What is done at this stage:

1. Tokenization is performed using nltk.word\_tokenize.
2. The text is converted to lowercase.
3. Stop words in English (prepositions, conjunctions, articles) are removed.
4. Tokens are filtered – characters that are not words are excluded.
5. Bigrams are formed.

The frequency of occurrence of each bigram in the entire text is calculated. The top 10 bigrams by frequency are selected, which are considered the main topics of this article. For example, these could be “global economic”, “military forces”, “energy crisis” and others. After this, a final file is formed with a list, which includes the found topics. For example:

"*topics*": [

"us dollar",

"members partners",

"new development",

"development bank",

"western countries",

"member countries",

"brics members",

"new members",

"brics partnership",

"financial institutions"

]

Let's move on to the next step.

**Thematic classification and tonality**

The purpose of this stage is to answer key questions:

* What do the news generally talk about?
* How do they talk about it – negatively, positively or neutrally?

At this stage, the main topic of each article was determined, after which the data was aggregated by country for the purpose of further analysis of the information received. In addition, the tone of the articles was determined.

The main topics by which the articles were classified were:

* finance;
* geopolitics;
* technology;
* trade;
* energy;
* military.

Each topic is represented by a list of keywords and their occurrence in each article is counted. An article can be assigned to several topics if they have scored the same number of “points”.

After this, the text sentiment is analyzed using the TextBlob library. This library is used for text processing. In particular, we were interested in analyzing the polarity of the text. If the polarity calculation showed a positive value, the text has a positive tonality, negative – with a negative tonality, and if 0, the text is neutral.

After this, the data is aggregated for all of each country and dictionaries are formed for each country with the calculated values – topic and tonality.

The next stage is analysis.

**Aggregation of article analysis results**

This stage is a logical continuation of the previous one. It allows you to turn the raw array of classification results obtained earlier into summary statistics, which are suitable for further visual analysis and interpretation of current geopolitical trends.

The goals of this stage are:

1. Calculating the total number of articles for each country.
2. Calculating the average sentiment of publications on a given topic for each country.
3. Determining the statistical distribution of topics (finance, geopolitics, technology, etc.) as a percentage of the total number of publications in the country.

Aggregation occurs as follows:

* A dictionary is created for each country with a count of:
  + the total number of articles,
  + a list of sentiments,
  + the number of mentions of each topic.
* In this case, the sentiment is calculated based on the average value, and the distribution of topics is normalized, that is, reduced to a share of the total number of articles.

After that, the result is saved in a separate file, where the country name and aggregated parameters are recorded for their further visualization and analysis.

Let's move on to the last stage - visualization.

**Visualization**

The goal of this stage is to create distribution graphs, bar charts and geomaps to obtain visual interpretations of the obtained aggregated values. This will make it easy to identify certain patterns in geopolitical topics, sentiment and media interest in the BRICS topic.

The initial data is the same - a dictionary with keys by country names and a dictionary value, which in turn contains keys reflecting the calculated value (the total number of news articles, average sentiment and a dictionary with the distribution of topics).

Libraries used at this stage:

1. Pandas is a standard library for working with tabular data.
2. Matplotlib.pyplot - for plotting graphs.
3. Seaborn - for improved visualization of graphs.
4. Geopandas is a library for processing and analyzing tabular data containing information about countries, as well as individual points, lines and polygons that can be displayed on maps.

So, first of all, we converted the data from the file to a standard DataFrame for pandas. After that, they started drawing graphs:

1. *Graph of the number of articles by country.*

This graph allows you to see how popular a given topic is in the media space of different countries. (Fig. 1) Since the news review was conducted from the moment of the request into the past, it is possible to estimate in which countries this topic was discussed more often, and in which less often.

A graph of articles

Description automatically generated

Fig. 1 – Number of articles per country on the topic “BRICS”

As you can see from this figure, only three countries have a heightened interest in BRICS – Canada, India and the USA. Great Britain also shows a high interest. However, other countries – Australia, China, Ireland, the Philippines, Pakistan and Singapore – are less interested in this topic. However, it is important to remember that at this stage we only downloaded English-language news, which gives a certain statistical “imprint” on the obtained values.

1. Graph of average news sentiment by country.

This graph shows the average emotional “coloring” of news articles (positive connotation, negative or neutral). The graph itself is shown in Figure 2.

A graph of a graph with different colored bars

Description automatically generated with medium confidence

Fig. 2 – Average sentiment per country

As we can see, the majority of the countries presented are inclined to a more positive coloring of news about BRICS. The exceptions are three countries - China, Great Britain and the USA. As for China, it is worth recalling that there is only one news item on this topic and therefore the sample is not representative enough to give any adequate and objective assessments. But the situation with Great Britain and the USA shows that in their media space BRICS is mainly described neutrally, but still with a slight negative coloring. This may be an element of the confrontation between NATO and the EU with BRICS, which is why we see such a coloring of news in these countries.

1. Stacked bar chart of topic distribution by country.

This chart shows composite bars that show the share of each news group in each country's news. You can see this chart in Figure 3.

A graph with different colored rectangles

Description automatically generated

Fig. 3 – Theme distribution per country

As you can see from this graph, some of the most popular and extensive topics are geopolitics and technology. The least popular topics are energy and military, which are not even represented in all countries. This shows the general trend of news articles related to the BRICS topic. That is, in the news where BRICS is present, they very often write about geopolitics for obvious reasons, and they also write quite often about technology, which may show the importance of technological exchange between the countries of the association and external countries interacting with this association. In third place in popularity is trade. This is also an equally important component of the information space, which is very closely related to geopolitics and exchange between countries. Let's move on to the next graph.

1. Geographical visualization for the number of articles.

This is a bit of an unconventional chart, as it is a map of the distribution of the number of articles by country. This will allow us to better understand which countries are more and less interested in the BRICS topic. You can see the map in Figure 4.

A map of the world with different countries/regions

Description automatically generated

Fig. 4 – Geographic coverage of news articles by country

In general, you can see the same results that were obtained earlier, but with an understanding of geography, which can be an important factor for additional analysis, for example, in a review of military conflicts or any local news of various kinds. Let's move on to the next graph.

1. Geographical distribution of news sentiment.

And the last graph being reviewed is the distribution of the news tone by country on a geographical map. You can see this graph in Figure 5.

A map of the world

Description automatically generated

Fig. 5 – Geographic coverage of news articles by country

This chart shows the average news sentiment by country, which can also be useful in analyzing how different countries feel about local events. The closer a country is to an event, the more likely the news will be oriented in one tone or another, depending on the type of event.

**Conclusions on the section**

So, we can draw the following conclusions on this section:

1. BRICS is a media political platform that is an important link in modern geopolitics and media space. Nevertheless, not that much attention of foreign mass media is paid to this alliance, making BRICS a promising research object rather than a popular one by now.
2. Using the developed algorithm, you can study the interest of different countries with a particular topic, in particular the topic of BRICS, look at the general orientation of the news - positive, negative, neutral - and also look at the dynamics of the thematic diversity of news articles.
3. The work shows that, using the possibilities of collecting open news information, it is possible to observe certain patterns in the behavior of the media using NLP and visualization, which can signal the current state of things in the media space around a particular topic.

The next stage is to check the operation of the algorithm on another topic.

**RUSSIA**

As a new keyword, it was decided to take the name of our country - Russia. There have always been a lot of news around our country, since we have a huge territory, a large army, we occupy an important place in global policy and trade, etc., but in addition to this in 2022, Russia joined the “Ukrainian conflict”, which still lasts and “crab” the world. Therefore, news about our country should be in abundance.

In addition, it was decided to modernize the algorithm - to translate articles from native languages of countries into English so as not to exclude them, but to add to the general analytical field.

Let's go through the stages of this section.

**Data collection**

This item did not change, the only thing was removing the filter in English, and of course the keyword with “Russia” was replaced.

**Parsing and machine translation**

This stage has undergone some changes.

In addition to obtaining the full text of the articles, we also completed the translation of them from the original language into English with the preservation of its original language in one of the dictionary keys. This allowed us to expand the list of countries and news that can be analyzed.

For the translation, the DeepTranslator library was used and, in particular, one of its classes - GoogleTranslator. It was the translated text of the article that was added to the final dictionary.

Thanks to this, we have expanded the list of countries participating in the study. If at the previous stage with BRICS we had 10 countries that we could study, now we have 38 countries. This makes the sample more representative and extensive.

**Text cleaning and segmentation**

This stage also did not undergo any significant changes:

1. Links, HTML tags, and double spaces, if any, were removed from the text.
2. The text was divided into sentences using nltk.sent\_tokenize.
3. Uninformative sentences were filtered out, i.e. those that consisted of a couple of words or were clearly “noisy”.
4. The cleaned articles were saved to files.

**Thematic and sentimental classification**

At this stage, pre-defined topics were also used, according to which news items were classified according to the type of text they contained. For each article, the probability of assignment to each topic was calculated.

In addition, the tonality of each news item was calculated according to the same criterion - positive, negative or neutral.

All aggregated parameters were saved in the corresponding variables for each news item.

**Analysis and visualization**

At this stage, the obtained results were visualized using the same matplotlib.pyplot and seaborn libraries.

The first graph is the distribution of the original languages of the articles that participated in the study. You can see this graph in Figure 6.

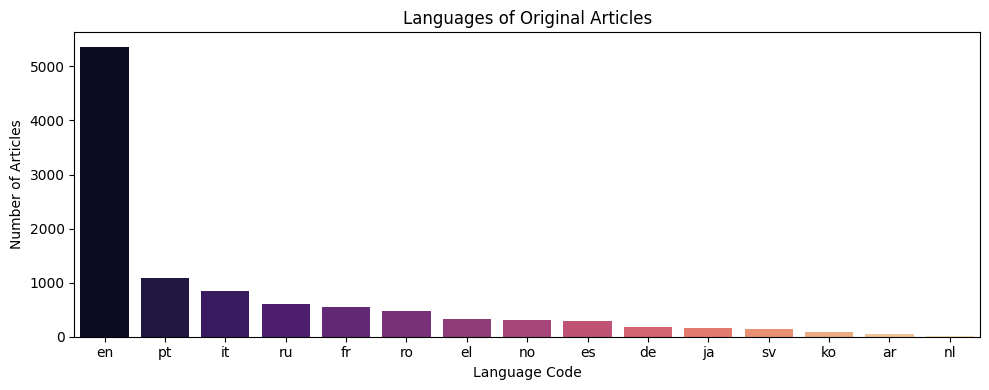


Fig. 6 – Distribution of languages of original articles

As you can see, the vast majority of articles were written in English, but still quite a lot of articles were written in Portuguese, Italian, Russian, etc.

This may indicate that authors of news articles are now mainly trying to stick to the “common field” and write in English, a universal language for almost the entire world.

Next, a table was created that shows how accurately the article texts were distributed by topic. In the table below, you can see the top 10 such news items with their accuracy of assignment to a particular topic.

| **Topic** | **Score** | **Title with Link** | **Lang** |
| --- | --- | --- | --- |
| Conflict | 0.998305 | [Israel-Politiker droht Russland im russischen Staats-TV](https://www.merkur.de/politik/israelischer-spitzenpolitiker-droht-im-russischen-staatsfernsehen-russland-wird-den-preis-dafuer-bezahlen-92592893.html) | de |
| Crime | 0.998159 | [G7 vai 'maximizar a pressão' se a Rússia rejeitar um cessar-fogo na Ucrânia](https://www.em.com.br/internacional/2025/05/7153875-g7-vai-maximizar-a-pressao-se-a-russia-rejeitar-um-cessar-fogo-na-ucrania.html) | pt |
| Conflict | 0.998070 | [The Guardian view on Russia sanctions: a brittle economy is Putin’s weakness](https://www.theguardian.com/commentisfree/2025/may/21/the-guardian-view-on-russia-sanctions-a-brittle-economy-is-putins-weakness) | en |
| Conflict | 0.997289 | [Oil touches $80 per barrel as sweeping sanctions against Russia rattle markets](https://finance.yahoo.com/news/oil-touches-80-per-barrel-as-sweeping-sanctions-against-russia-rattle-markets-162612572.html) | en |
| Economy | 0.996606 | [The Guardian view on Russia sanctions: a brittle economy is Putin’s weakness](https://www.theguardian.com/commentisfree/2025/may/21/the-guardian-view-on-russia-sanctions-a-brittle-economy-is-putins-weakness) | en |
| Economy | 0.996394 | [To end the war in Ukraine, make Beijing's bankrolling of Russia's war machine hurt](https://nypost.com/2025/05/19/opinion/to-end-the-war-in-ukraine-make-beijings-bankrolling-of-russias-war-machine-hurt/) | en |
| Diplomacy | 0.996313 | [GOP congressman confirms Hegseth ordered pause in cyber operations against Russia, despite Pentagon denial](https://www.cbsnews.com/news/gop-congressman-don-bacon-hegseth-ordered-pause-cyber-operations-against-russia/) | en |
| Technology | 0.996281 | [Google suspende la monetización de medios estatales rusos en sus plataformas](https://larepublica.pe/economia/2022/02/27/guerra-entre-rusia-y-ucrania-google-suspende-la-monetizacion-de-medios-estatales-rusos-en-sus-plataformas/) | es |
| Conflict | 0.995994 | [Anthony Albanese and Penny Wong slam Russia over 13-year sentence for captured fighter](https://7news.com.au/news/anthony-albanese-and-penny-wong-slam-russia-over-13-year-sentence-for-captured-fighter-c-18727462) | en |
| Conflict | 0.995441 | [Il avait suggéré de noyer des enfants ukrainiens, un journaliste de la chaîne Russia Today suspendu](https://www.ladepeche.fr/2022/10/24/il-avait-suggere-de-noyer-des-enfants-ukrainiens-un-journaliste-de-la-chaine-russia-today-suspendu-10759853.php) | fr |

From the short descriptions of the articles you can see that the classification is done correctly.

After that, two lists of countries were created - friendly and unfriendly towards Russia. And on the basis of these lists, sentiment graphs were created (Figures 7 and 8). In these graphs you can see the attitude of friendly and unfriendly countries towards Russia.

As it can be seen from the graphs, both lists contain countries that are mostly (or even completely) negative about Russia. At the same time, both groups do not contain a single country that is completely positive about Russia. This data can be used in further research on our country's relations with other countries, including those which are considered "friendly".

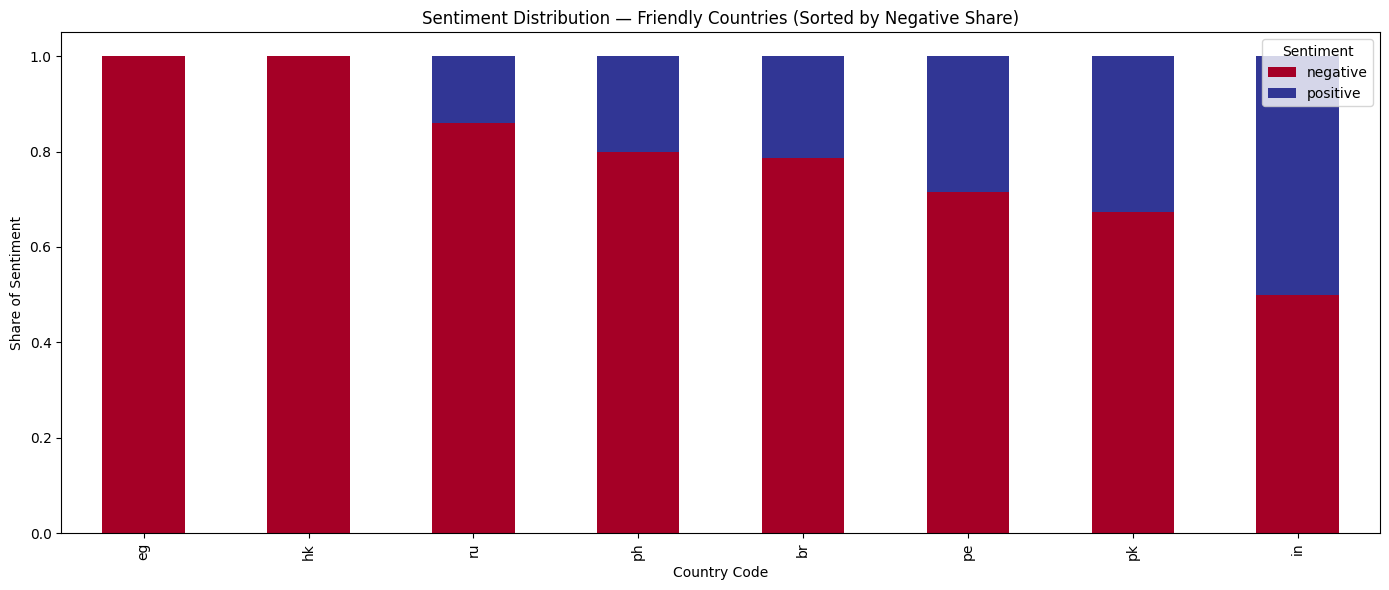


Fig. 7 – Sentiment Distribution by friendly countries

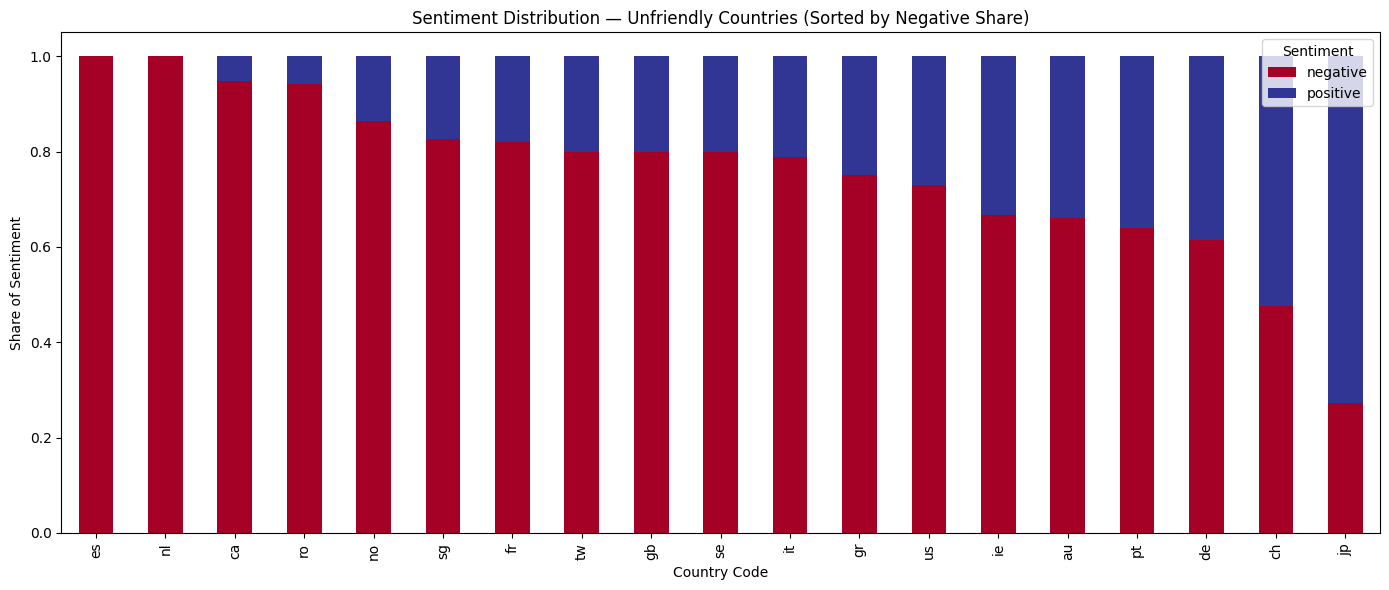


Fig. 8 – Sentiment Distribution by unfriendly countries

The bar chart above shows very captivating in terms of geopolitical trends results.

Japan stands out with a more balanced sentiment distribution, defying the initial interpretation, even though the general trend across these nations indicates a preponderance of negative sentiment toward Russia. Most nations have a negative sentiment share of between 0.6 and 1.0, but Japan has a much higher positive sentiment share—nearly 0.7—which suggests a more complex and even positive media representation and may serve as a sign of improving relations between Russia and Japan. And in fact, even after conduction this analysis an important event occurred: Russian President V. Putin received in the Kremlin the widow of former Japanese Prime Minister Shinzo Abe, Akie Abe[[4]](#footnote-3). By some experts it is considered from the side of Tokyo as a hint to resume relations between two countries[[5]](#footnote-4).

The European nations with the highest percentages of negative sentiment (0.9–1.0) and little to no positive sentiment are Romania (ro), France (fr), Switzerland (ch), Spain (es), Germany (de), and Italy (it). This illustrates how the European Union has united against Russia's actions in Ukraine, especially since the conflict escalated in 2022. Romania, according to the graph, is a “leader” in terms of negative sentiment about Russia – the country’s authorities state that Russia poses special danger to Romania due to its close to Ukraine location. Such perception especially escalated because of recent elections in Romania, during which Romanian media accused[[6]](#footnote-5) Russian media interfering into its internal policy [[7]](#footnote-6)

It is also worth noting an important aspect of these graphs. The tonality was calculated based on the text of the entire news item, i.e. it is important to understand that “negative” and “positive” news items are not always reflected in the negative and positive attitude of countries in this case towards Russia. They show that the color of the news item itself is negative or positive. For example, this could be news about a terrorist attack on Russian territory or about the start of an international forum on Russian territory, which may not have a direct relationship between the country where the news item was made and Russia itself.

Taking all this into account, this section considered a new keyword - "Russia". Some changes and additions were made to the algorithm, which showed that it is possible to adapt the mechanism to the specific requests of those who use this mechanism.

The following comparative conclusions can be made for this and the previous section:

1. For BRICS, many countries "returned" a very small amount of material, in some there was none at all, which indicates a narrow interest in this topic by now. However, it can be related to the fact that the organization is relatively young and yet there are no integrational processes among BRICS countries, as well as lack of institutionalization. The same cannot be said about the word "Russia", for which there was quite a lot of news in each country, and therefore a larger sample size for analysis. Russia is a more stable and large-scale geopolitical topic compared to BRICS.
2. Translating articles from the original language into English was a right decision, since we were thus able to significantly increase the sample for analysis.

**Conclusions on the empirical part**

A full cycle of automatic news analysis has been successfully implemented with the ability to specify certain steps depending on the purpose of the analysis:

1. Collection by keywords using the GNews API.
2. Downloading the full text of articles.
3. Cleaning, marking, segmentation, classification by topic.
4. Sentiment analysis.
5. Visualization (graphs, maps).

The mechanism has shown good performance on real articles taken by keywords. Thanks to the created mechanism, it is possible to analyze the current media space and its components, which can help in making economic, political and other decisions.

As further prospects for improving the mechanism, the following points can be highlighted:

1. Changing the standard BERT language model to a multilingual one, so that there is no need to translate articles into English, since this still introduces a certain noise into the data and the quality of further classification and tonality analysis is lost.
2. Conduct a deeper analysis of article texts, take into account metaphors, rhetorical constructions of different languages, which are not highlighted in a special way by conventional language models, but have an additional color in natural language.
3. Temporal analysis: the ability to see how trends have changed over time. For example, regarding Russia, what was the tonality of the news before the conflict in Ukraine, at its very beginning and now.
4. Improve automatic processing by introducing asynchronous computations, parallelization, and also generally optimize the code by replacing more “heavyweight” code constructs with lighter and faster ones.

# **Conclusion**

This master's thesis was devoted to the development of mechanisms for identifying current geopolitical trends through media content analysis using modern data analysis methods, including natural language processing (NLP) and tonality analysis. The research aimed to create a universal algorithm that could serve as a tool for analyzing large amounts of news data, identifying key topics and emotional coloring of publications, as well as providing visualized results for interpreting geopolitical trends. As part of the work, the following research tasks were set and consistently solved, which made it possible to achieve the goals set and contribute to the study of this area.

1. Conducting a literature review on existing media-based approaches to analyzing geopolitical trends.

A review of the literature confirmed that media analysis is a powerful tool for identifying geopolitical trends, but existing approaches are often limited by isolated consideration of events, insufficient versatility of algorithms, and language barriers. Studies on the use of machine learning and NLP have shown their effectiveness in predicting conflicts and analyzing public opinion, but there are not enough such works in the Russian scientific base, which underlines the relevance and novelty of this study. An analysis of the literature also revealed the need to integrate quantitative and qualitative methods for a deeper understanding of geopolitical narratives, which became the basis for the development of the proposed algorithm.

2. Formulation of cases for testing the algorithm.

Two key words were chosen as cases: "BRICS" and "Russia". The choice of BRICS was due to its importance as an alternative global governance platform and the resonance caused by the expansion of the organization in 2023. Russia, in turn, is a large-scale geopolitical entity, especially in the context of the conflict in Ukraine, which began in 2022 and continues in 2025. These cases allowed us to test the algorithm at various levels: from analyzing narratives around economic and political cooperation (BRICS) to studying the perception of a country , especially during the time of the global conflict (Russia).

3. Collection and preprocessing of news articles using the GNews API. Data collection was successfully carried out using the GNews API. For the BRICS case, English—language articles from 10 countries were collected, and for the Russia case, from 38 countries, including translations of texts from native languages into English. Preprocessing included cleaning the texts from HTML tags, segmentation into sentences, and removal of uninformative data, which provided a high-quality database for subsequent analysis.

4. The use of natural language processing techniques to identify key themes and trajectories of tonality.

The use of NLP techniques, such as bigram extraction and tonality analysis using the TextBlob library, made it possible to identify the dominant themes and emotional coloring of the news. For BRICS, the main topics were geopolitics and technology, reflecting the media's focus on the strategic importance of the organization. The analysis of the key topics showed a predominantly positive or neutral perception of BRICS in most countries, with the exception of the United States and Great Britain, where the key was slightly negative, probably due to competition with Western structures such as NATO. For Russia, the themes of conflict predominated, which is related to the war in Ukraine, and the tone was predominantly negative in both friendly and unfriendly countries, which underlines the global impact of the conflict.

5. Comparison of media narratives in different countries.

A comparative analysis revealed significant differences in the perception of BRICS and Russia. For BRICS, the countries of North America (Canada, USA) and South Asia (India) showed the greatest interest, while in other countries, such as China and Singapore, interest was minimal, which may be due to the linguistic limitations of the sample.

In the case of Russia, the negative perception was universal, but in unfriendly countries such as Romania and France, the share of negative sentiment reached 90-100%, while in Japan there was a more balanced and even positive picture.

6. Assessment of the strengths and weaknesses of the applied analytical methods.

The developed algorithm has demonstrated high efficiency in processing large amounts of data and the accuracy of topic classification (for example, the accuracy of the distribution of articles by topic was up to 99.8%). The use of open models such as TextBlob has ensured transparency and reproducibility of the results. However, the limitations are related to the language bias (the predominant adaptation of tools to English), differences in the editorial policy of media sources, and the difficulty of taking into account the cultural context. The translation of texts from native languages into English for the Russia case also introduced some noise into the data, which requires the introduction of multilingual models in the future.

7. Assessment of the predictive potential of the developed mechanism.

The mechanism has shown high predictive potential, allowing it to identify current trends and potential stress points. For example, the negative tone towards Russia in unfriendly countries indicates the country's continued isolation in the international arena, which can be used to predict further sanctions or escalation of conflicts. For BRICS, the analysis revealed a growing interest in technological cooperation, which may signal future initiatives in this area. Data visualization (graphs, maps) has facilitated the interpretation of the results, making them applicable to research and government agencies.

8. Formulation of the algorithm and provision of conclusions.

The result of the work was a step-by-step algorithm that includes data collection, text extraction, preprocessing, thematic and tonal analysis, as well as visualization. This algorithm can be adapted to various geopolitical cases and scaled for real-time, which makes it a valuable tool for analyzing the media space. The case studies showed that BRICS is perceived as a significant platform for geopolitical and technological cooperation, while Russia is under strong negative pressure in the global media, which requires further study of the dynamics of relations with other countries.

Additional findings and significance of the study

The empirical part of the work confirmed that the developed mechanism is able to effectively analyze media content, identifying both general trends and country-specific features. For example, geographical visualization showed the concentration of interest in BRICS in North America and India, and tonality analysis revealed differences in perception between countries. The research has practical significance, as the proposed algorithm can be used to create an early warning system about geopolitical risks, as well as to support strategic decision-making in conditions of political instability.

In conclusion, this work has contributed to the development of media analysis methodology for geopolitical research by offering a universal and transparent approach. Prospects for further research include the introduction of multilingual models, time analysis to track trend dynamics, and the development of a web interface for interacting with the mechanism in a real-time, which will expand its applicability in real-world conditions.

As a result of this thesis, an algorithm was created that allows analyzing news articles by keywords and making conclusions based on graphs and metrics of sentiment and frequency distribution of topics.

Two keywords were considered for selecting news articles - "BRICS" and "Russia". Different methods of processing, preparing and analyzing articles were carried out, and corresponding conclusions were made.

The following can be highlighted as prospects for the development of the project:

1. Transition to forecasting through time series analysis or recurrent neural networks.
2. Deepening topic modeling using LDA, BERTopic and other methods, moving away from the bigram method.
3. Integrate multilingual models to level out "misses" in translation.
4. Creation of an "alarm" system if a local event has occurred, so that it can be monitored through the media in real time and decisions can be made as early as possible.
5. Creation of a web interface for interaction with the mechanism.

# **Bibliography**

1. Albanese F., Pinto S., Semeshenko V., Balenzuela P. Analyzing Mass Media Influence Using Natural Language Processing and Time Series Analysis [online] // *Journal of Physics: Complexity*. 2020. Vol. 1. Article 025005. Available at: <https://doi.org/10.1088/2632-072X/ab8cb9> (accessed: 13.04.2025).
2. Anwar A., Iltaf N., Iqbal J. Social Media Analysis to Gauge Geopolitical Sentiment: A Case Study of the 2021 US Withdrawal from Afghanistan // Frontiers in Big Data. 2022. Vol. 5. Art. 886362. doi:10.3389/fdata.2022.886362
3. Arradon, I. Crisis Early Warning and Horizon Scanning in Today’s World [Electronic resource] / Isabelle Arradon // International Crisis Group. – URL: https://www.crisisgroup.org/global/crisis-early-warning-and-horizon-scanning-todays-world (accessed: 03.05.2025).
4. Benoit K. Text as Data: An Overview [online] // The SAGE Handbook of Research Methods in Political Science and International Relations. London: SAGE Publications Ltd., 2020. P. 461-497. Available at: https://methods.sagepub.com/book/research-methods-in-political-science-and-international-relations/i4365.xml (accessed: 22.03.2025).
5. Bekaert, G., Harvey, C. R., Lundblad, C. T., & Siegel,S. 2014. Political risk spreads. Journal ofInternational Business Studies, 45(4): 471–493.
6. Boydstun, A. E. How Media Shapes Politics [Electronic resource] / A. E. Boydstun // UCI GCC Podcast. – 2022. – URL: https://ucigcc.org/podcast/how-media-shapes-politics/ (accessed 30.03.2025).
7. Brennen, B. S. Qualitative Research Methods for Media Studies [Text] / B. S. Brennen. – 3rd ed. – New York ; London : Routledge, 2022. – 296 p. – ISBN 978-0-367-64153-5 (hbk), 978-0-367-64150-4 (pbk), 978-1-003-12238-8 (ebk).
8. Burns, J. C., Kelsey, T., & Donovan, C. (2025). Extracting emerging events from social media: X/Twitter and the multilingual analysis of emerging geopolitical topics in near real time. Journal of Social Media Research, 2(1), 50-70. doi:10.29329/isomer.14
9. Caldara, D., Iacoviello, M. Measuring Geopolitical Risk // Working Paper. – 2022. – 76 p. – URL: https://www.matteoiacoviello.com/gpr\_files/GPR\_PAPER.pdf (accessed 23.03.2025).
10. Democratic Erosion. Russia Fueling Democratic Struggles in Romania Through the Media [online] // Democratic Erosion. 2025. Available at: https://democratic-erosion.org/2025/04/24/russia-fueling-democratic-struggles-in-romania-through-the-media/ (accessed: 20.05.2025).
11. Denny, M. J., and Spirling, A. (2018). Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it. *Polit. Anal*. 26, 168–189. doi: 10.1017/pan.2017.44
12. De Slegte, J., Van Droogenbroeck, F., Spruyt, B., Verboven, S., & Ginis, V. (2024). The Use of Machine Learning Methods in Political Science: An In-Depth Literature Review. *Political Studies Review*, *0*(0).
13. Diaz, F., Gamon, M., Hofman, J. M., Kiciman, E., and Rothschild, D. (2016). Online and social media data as an imperfect continuous panel survey. *PLoS ONE* 11, e0145406. doi: 10.1371/journal.pone.0145406
14. DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data Soc*. 2:2053951715602908. doi: 10.1177/2053951715602908
15. Dubois, E., & Blank, G. (2018). The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media. Information, Communication & Society, 21(5), 729–745. doi:10.1080/1369118X.2018.1428656
16. Eberl, J.-M., Jacobi, C., Schlögl, S. Exploring news content complexity using network analysis // International Communication Association (ICA), Social and Semantic Networks in Communication Research Preconference, Seattle, May 22, 2014. – 2014. – P. 1–33.
17. Fuelling the fire: Russia's role in Romania's democratic struggles [online] // Kew Royal Botanic Gardens. 2025. Available at: <https://www.kew.org.pl/en/2025/02/14/fuelling-the-fire-russias-role-in-romanias-democratic-struggles/> (accessed: 20.05.2025).
18. GDELT Project [Electronic resource] // Global Database of Events, Language, and Tone. – URL: https://www.gdeltproject.org (accessed: 24.05.2025)
19. Geopolitical Forecast // Geopolitical Futures. – 2022. – September 13. – URL: https://geopoliticalfutures.com/forecasting/ (accessed 23.02.2025).
20. GNews Documentation (v4) [Electronic resource] // GNews API. – URL: <https://gnews.io/docs/v4#introduction> (accessed: 28.03.2025).
21. Google Trends [Electronic resource]. – URL: <https://trends.google.com/trends/> (accessed: 24.04.2025)
22. Graff A., Chousou-Polydouri N., Inman D., et al. Curating Global Datasets of Structural Linguistic Features for Independence // Scientific Data. 2025. Vol. 12. Art. 106. doi:10.1038/s41597-024-04319-4
23. Halterman,A, Bagozzi, B., Beger,A.,Schrodt,P. Scraborough, G (2023). PLOVER and POLECAT: A New Political Event Ontology and Dataset. Arxiv Social Science.
24. Hamidreza Mostafaei, Shirin Kordnoori, Mohammadmohsen Ostadrahimi, Saeed Seyed Agha Banihashemi, “Applications of artificial intelligence in global diplomacy: A review of research and practical models”, Sustainable Futures, Volume 9, 2025, ISSN 2666-1888, <https://doi.org/10.1016/j.sftr.2025.100486>.
25. Hinck, R. (Ed.) Geopolitical Visions in Russian Media [Electronic resource] // NSI Team. - URL: https://nsiteam.com/social/geopolitical-visions-in-russian-media/ (accessed: 29.03.2025).
26. Hoffmann, M., Santos, F. G., Neumayer, C., &Mercea, D. 2022. Lifting the veil on the use of bigdata news repositories: A documentation and criticaldiscussion of a protest event analysis.Communication Methods and Measures, 16(4):283–302.
27. How are geopolitical risks affecting the world economy? // Economics Observatory. – 2024. – March 1. – URL: https://www.economicsobservatory.com/how-are-geopolitical-risks-affecting-the-world-economy (accessed 20.03.2025).
28. Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets [Text] / A. S. Imran, S. M. Daudpota, Z. Kastrati, R. Batra // IEEE Access. – 2020. – Vol. 8. – P. 181074–181090. – DOI: 10.1109/ACCESS.2020.3027350.
29. Integrated Crisis Early Warning System (ICEWS) [Electronic resource] // Lockheed Martin. - URL: https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html (accessed: 25.03.2025)
30. Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Zhao, T. SMART: Robust and Efficient Fine-Tuning for Pre-trained Natural Language Models through Principled Regularized Optimization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 2177–2190.
31. Kukštaitė K. Triadic Relationship Between Social Movement, News Media, and Geopolitics: Government Affiliated Transnational Media and its Coverage of 2020–2021 Belarusian Protests // Stockholm University, Department of Media Studies, 2021. – 83 p.
32. Kupfer, A., & Zorn, J. 2020. A language-independent measurement of economic policy uncertainty in Eastern European Countries. Emerging Markets Finance and Trade, 56(5): 1166–1180.
33. Lee, J., Cha, J., Park, E. Data-driven approaches into political orientation and news outlet discrimination: The case of news articles in South Korea // Telematics and Informatics. – 2023. – Vol. 85. – Article No. 102066. – Available at: https://www.sciencedirect.com/science/article/abs/pii/S0736585323001302 (accessed 30.03.2025).
34. Li, C.; Xue, R.; Lin, C.; Fan, W.; Han, X. CUCPLUS at CheckThat! 2023: Text Combination and Regularized Adversarial Training for News Media Factuality Evaluation. In Proceedings of the CLEF 2023: Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 18–21 September 2023. [Google Scholar]
35. Lin, Chichen, Yongbin Wang, Chenxin Li, Weijian Fan, Junhui Xu, and Qi Wang. 2024. "A Cross-Lingual Media Profiling Model for Detecting Factuality and Political Bias" Applied Sciences 14, no. 21: 9837. <https://doi.org/10.3390/app14219837>
36. Liu Y., Li D., Feng J. Incorporating Cross-Linguistic and Time-Based Dimensions to Critical Metaphor Analysis: A Specialised Hands-On Analytical Approach // Critical Discourse Studies. 2024. doi:10.1080/17405904.2024.2393388
37. Machine Learning for Peace // Observatory of Public Sector Innovation. 2023. URL: https://oecd-opsi.org/innovations/machine-learning-for-peace/ (accessed 20.04.2025).
38. Mäntylä, M. V., Graziotin, D., and Kuutila, M. (2018). The evolution of sentiment analysis-a review of research topics, venues, and top cited papers. *Comput. Sci. Rev*. 27, 16–32. doi: 10.1016/j.cosrev.2017.10.002
39. Mickoleit, A. (2014). Social media use by governments: A policy primer to discuss trends, identify policy opportunities and guide decision makers. OECD Working Papers on Public Governance, No. 26, 70 p. https://dx.doi.org/10.1787/5jxrcmghmk0s-en
40. Monitoring Geopolitical Risk: Strategies and Resources for the Private Sector Intelligence // Feedly TI Essentials. 2024. URL: https://feedly.com/ti-essentials/posts/monitoring-geopolitical-risk-strategies-and-resources-for-the-private-sector-intelligence (accessed 25.04.2025).
41. Moralwar S. B., Deshmukh S. N. Different Approaches of Sentiment Analysis [Electronic resource] // International Journal of Computer Sciences and Engineering. – 2015. – Vol. 3, no. 3. – Pp. 160–165. – URL: [www.ijcseonline.org](http://www.ijcseonline.org) (accessed: 24.05.2025).
42. Mejias U.A., Vokuev N.E. Disinformation and the Media: The Case of Russia and Ukraine // Media, Culture & Society. – 2017. – Vol. 39, No. 7. – Pp. 1027-1042. – DOI: <https://doi.org/10.1177/0163443716686672>. – (Accessed: 24.04.2025).
43. Mueller, H., & Rauh, C. (2013). Machine Learning and Conflict Prediction: A Use Case. Stability: International Journal of Security and Development, 2(3), Art. 56. doi:10.5334/sta.cr
44. Ortega‑Martorell, S., Bellfield, R. A. A., Harrison, S., Dyke, D., Williams, N., Olier, I. Mapping the global free expression landscape using machine learning // SN Applied Sciences. – 2023. – Vol. 5. – Article 354. – Available at: https://doi.org/10.1007/s42452-023-05554-x (accessed 30.03.2025).
45. Ortiz, Áort., Rodrigo, T., Sicilia, J.  
    *BBVA Research Geopolitics Monitor: Tracking Geopolitical Sentiment and Events Using Natural Language Techniques* [Electronic resource] // SUERF Policy Notes and Briefs. - 2023. - No. 305. - URL: <https://www.suerf.org/publications/suerf-policy-notes-and-briefs/the-bbva-research-geopolitics-monitor-tracking-geopolitical-sentiment-and-events-using-natural-language-techniques/> (accessed: 28.03.2025).
46. Pellegrino, A. Research Trends in Social Media Role in Political Campaigns: A Bibliometric Analysis / A. Pellegrino // Review of Communication Research. – 2023. – Vol. 11. – P. 150–171. – DOI: 10.5680/RCR.V11.6.
47. Puhr, H., & Kupfer, A. 2023. Media in the Geopolitical Crossfire: Identification and Novel. Data Sources for IB Research. AIB Insights.
48. Puschmann, C., and Powell, A. (2018). Turning words into consumer preferences: how sentiment analysis is framed in research and the news media. *Soc. Media Soc*. 4:2056305118797724. doi: 10.1177/2056305118797724.
49. Radaideh A., Dweiri F. Sentiment Analysis Predictions in Digital Media Content using NLP Techniques [online] // *International Journal of Advanced Computer Science & Applications*. 2023. Vol. 14. No. 11. P. 283. ISSN 2158-107X. Available at: <https://doi.org/10.14569/IJACSA.2023.0141128> (accessed: 21.04.2025).
50. Ramakrishnan, N., Butler, P., Self, N., Khandpur, R.P., Saraf, P., Wang, W. Attention-Based LSTM for Predictive Modeling of Societal Events Using Sequential Text Data [Electronic resource] // Frontiers in Artificial Intelligence. - 2022. - Vol. 5. - Article 893875. - URL: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2022.893875/full (accessed: 03.05.2025)
51. Ross, A. R. N., Chadwick, A., & Vaccari, C. (2021). Digital Media and the Proliferation of Public Opinion Cues Online: Biases and Vulnerabilities in the New Attention Economy. In Morrison, J., Birks, J., & Berry, D. M. (eds), The Routledge Companion to Political Journalism. Routledge. doi:10.4324/9780429284571
52. Saraisky, N. G. (2015). Analyzing public discourse: Using media content analysis to understand the policy process. Current Issues in Comparative Education, 18(1), 26–41.
53. Schilling E., Munnes S., Harsch C., Knobloch M., Vogel J. S., Hipp L. Examining Sentiment in Complex Texts: A Comparison of Different Computational Approaches [online] // Frontiers in Big Data. 2022. Vol. 5. Article 886362. Section: Data Science. Available at: https://doi.org/10.3389/fdata.2022.886362 (accessed: 24.04.2025).
54. Shukla, D., Unger, S. Sentiment Analysis of International Relations with Artificial Intelligence [Electronic resource] // *Athens Journal of Sciences*. – 2022. – Vol. 9. – P. 1–16.  
    URL: <https://doi.org/10.30958/ajs.X-Y-Z> (accessed: 25.03.2025).
55. Stone T. Understanding Geopolitical Risk: How Data Intelligence Empowers Market Understanding // Finance Derivative. 2024. URL: https://www.financederivative.com/understanding-geopolitical-risk-how-data-intelligence-empowers-market-understanding/ (accessed 25.05.2025).
56. Strömberg, D. Media and Politics: discussion paper / David Strömberg. – London : Centre for Economic Policy Research, 2015. – 45 p. – (Discussion Paper Series; No. 10426). – Available at: https://cepr.org/active/publications/discussion\_papers/dp.php?dpno=10426 (accessed 30.03.2025).
57. Subekti, D., Nurmandi, A., & Mutiarin, D. (2022). Mapping publication trend of political parties campaign in social media: A bibliometric analysis. Journal of Political Marketing. doi:10.1080/15377857.2022.2104424.
58. The Kremlin. Meeting with permanent members of the Security Council [online] // Official website of the President of Russia. 2024. Available at: http://en.kremlin.ru/events/president/news/77060 (accessed: 30.05.2025).
59. Trubowitz P., Watanabe K. The Geopolitical Threat Index: A Text-Based Computational Approach to Identifying Foreign Threats // International Studies Quarterly. 2021. Vol. 65, No. 3. P. 852–865. doi:10.1093/isq/sqab029
60. University of Washington. Where Media and Politics Meet [Electronic resource] // Arts & Sciences News. – 2023. – URL: https://artsci.washington.edu/news/2023-10/where-media-and-politics-meet (accessed 30.03.2025).
61. Using OSINT in geopolitical assessment: a practical guide // Authentic8. – 2024. – December 10. – URL: https://www.authentic8.com/blog/osint-geopolitical-assessment (accessed 19.03.2025).
62. Vijesti. (2024). How Russia influenced the elections in Romania. Vijesti.me; Vijesti.https://en.vijesti.me/world-a/evropa/736807/how-Russia-influenced-the-elections-in-Romania
63. Vzglyad. [Article Title in Original Russian] [online] // Vzglyad.ru. 2025. Available at: https://vz.ru/news/2025/5/30/1335519.html (accessed: 30.05.2025).
64. Wang, H., Wang, X. Sentiment analysis of tweets and government translations: Assessing China’s post-COVID-19 landscape for signs of withering or booming // Global Media and China. – 2023. – Vol. 8, No. 2. – pp. 213–230. – URL: https://journals.sagepub.com/doi/full/10.1177/20594364231181745 (accessed 24.04.2025).
65. Wan X., Yang J., Wang Y., Zhang Q. Mining the Voice of the People: Media as a Source for Political Insight Extraction // Knowledge-Based Systems. 2017. Vol. 129. P. 1–13. doi: 10.1016/j.knosys.2017.04.017
66. Ward, M. D., Metternich, N. W., Dorff, C. L., Gallop, M., Hollenbach, F. M., Schultz, A., Weschle, S. Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction // International Studies Review. – 2013. – Vol. 15, No. 4. – pp. 473–490. – DOI: 10.1111/misr.12072.
67. Weijters B., Puntoni S., Baumgartner H. Methodological Issues in Cross-Linguistic and Multilingual Advertising Research [Electronic resource] // Lancaster University EPrints. – 2019. – URL: <https://eprints.lancs.ac.uk/id/eprint/131523/1/Linguistic_issues_in_advertising_research_plain_text.pdf>(accessed: 24.04.2025).
68. Windsor L. C. Bias in Text Analysis for International Relations Research // Global Studies Quarterly. 2022. Vol. 2, No. 3. P. 1–13. doi:10.1093/isagsq/ksac021
69. Zamaruieva, I., Lienkov, S., Babich, O., Shevchenko, A., Khlaponin, Y., Bernaz, N. Analytical approaches to news content processing during the war in Ukraine in opposing geopolitical alliances mass media // In: Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2023), April 20–21, 2023, Kharkiv, Ukraine. CEUR Workshop Proceedings, Vol. 3461. P. 241–270. URL: http://ceur-ws.org/Vol-3461/paper46.pdf
70. Zürn, M. Geopolitics, Economics, and the Future of Global Order // Oxford Research Encyclopedia of Politics. – 2019. – URL: https://oxfordre.com/politics/display/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-957 (accessed 19.03.2025).

1. GDELT Project [Electronic resource] // Global Database of Events, Language, and Tone. – URL: https://www.gdeltproject.org (accessed: 24.05.2025) [↑](#footnote-ref-0)
2. Google Trends [Electronic resource]. – URL: https://trends.google.com/trends/ (accessed: 24.04.2025) [↑](#footnote-ref-1)
3. Google Trends [Electronic resource]. – URL: <https://trends.google.com/trends/> (accessed: 24.04.2025) [↑](#footnote-ref-2)
4. The Kremlin. Meeting with permanent members of the Security Council [online] // Official website of the President of Russia. 2024. Available at: http://en.kremlin.ru/events/president/news/77060 (accessed: 30.05.2025). [↑](#footnote-ref-3)
5. Vzglyad. [Article Title in Original Russian] [online] // Vzglyad.ru. 2025. Available at: https://vz.ru/news/2025/5/30/1335519.html (accessed: 30.05.2025). [↑](#footnote-ref-4)
6. Vijesti. (2024). How Russia influenced the elections in Romania. Vijesti.me; Vijesti.https://en.vijesti.me/world-a/evropa/736807/how-Russia-influenced-the-elections-in-Romania [↑](#footnote-ref-5)
7. Democratic Erosion. Russia Fueling Democratic Struggles in Romania Through the Media [online] // Democratic Erosion. 2025. Available at: <https://democratic-erosion.org/2025/04/24/russia-fueling-democratic-struggles-in-romania-through-the-media/> (accessed: 20.05.2025). [↑](#footnote-ref-6)