DOI: 10.14515/monitoring.2026.4.3147

I. D. Petrov

POLARIZED WORDS, POLARIZED WORLDS: A COMPUTATIONAL COMPARATIVE ANALYSIS OF PRO- AND ANTI-VACCINE DISCOURSE ON VK

For citation:
Petrov I. D. (2026) Polarized Words, Polarized Worlds: A Computational Comparative Analysis of Pro- and Anti-Vaccine Discourse on VK. *Monitoring of Public Opinion: Economic and Social Changes.* No. 4. P. 50–71. https://www.doi.org/10.14515/monitoring.2026.4.3147.
Правильная ссылка на статью:
Петров И. Д. Поляризованные слова, поляризованные миры: сравнительный компьютерный анализ про- и антивакцинного дискурса во «ВКонтакте» // Мониторинг общественного мнения: экономические и социальные перемены. 2026. № 4. С. 50—71. https://www.doi.org/10.14515/monitoring.2026.4.3147. (In Eng.)

Получено: 18.10.2025. Принято к публикации: 28.04.2026.

DOI: 10.14515/monitoring.2026.4.3147

POLARIZED WORDS, POLARIZED WORLDS: A COMPUTATIONAL COMPARATIVE ANALYSIS OF PRO- AND ANTI-VACCINE DISCOURSE ON VK

Igor D. PETROV[1] *— Graduate student, Department of Sociology*
E-MAIL: idpetrov@hse.ru
https://orcid.org/0000-0002-8465-8728

[1] HSE University, St. Petersburg, Russia

Abstract. This study aims to compare the emotional and thematic differences in pro-vaccination and anti-vaccination discourse on the Russian social network VKontakte (VK), using Language Expectancy Theory (LET) as an interpretive lens. A computational approach was applied to analyze 18,239 posts from 229 anti-vaccination and 84 pro-vaccination communities, utilizing dictionary-based sentiment analysis and structural topic modeling. The results revealed that anti-vaccination content is characterized by more intense use of emotional and negative language, whereas pro-vaccination communities exhibit a more neutral and rational tone. Thematically, anti-vaccination discourse more frequently focuses on children, family, and conspiracy theories, while pro-vaccination discourse emphasizes science, research, and public health efficacy. However, the emotional tone of posts was not a significant predictor of user engagement. The study identifies systematic differences in the linguistic profiles of pro- and anti-vaccination communities, consistent with expectations that LET would predict distinct audience groups. However, since this study did not directly measure persuasiveness or audience expectations, these theory-informed interpretations remain tentative and require further empirical validation.

ПОЛЯРИЗОВАННЫЕ СЛОВА, ПОЛЯРИЗОВАННЫЕ МИРЫ: СРАВНИТЕЛЬНЫЙ КОМПЬЮТЕРНЫЙ АНАЛИЗ ПРО- И АНТИВАКЦИННОГО ДИСКУРСА ВО «ВКОНТАКТЕ»

ПЕТРОВ Игорь Дмитриевич — аспирант Департамента социологии, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, Россия
E-MAIL: idpetrov@hse.ru
https://orcid.org/0000-0002-8465-8728

Аннотация. В работе проводится сравнительный анализ эмоциональных и тематических различий в дискурсе провакцинных и антивакцинных сообществ в российской социальной сети «ВКонтакте». В качестве интерпретативной рамки используется теория языковых ожиданий. Для анализа 18 239 постов из 229 антивакцинных и 84 провакцинных сообществ применялись методы обработки естественного языка: анализ тональности на основе словарей и структурное тематическое моделирование. Результаты показали, что антивакцинный контент характеризуется более интенсивным использованием эмоционально насыщенной и негативной лексики, в то время как провакцинные сообщества придерживаются более нейтрального и рационального тона. Тематически антивакцинный дискурс чаще фокусируется на детях, семье и теориях заговора, тогда как провакцинный — на науке, исследованиях и эффективности общественного здравоохранения. При этом эмоциональность постов не оказала значимого влияния на вовлеченность пользователей. Исследование выявляет систематические различия в лингвистических профилях провакцинных и антивакцинных сообществ, которые согласуются с ожиданиями, предсказываемыми теорией языковых ожиданий для различных аудиторных групп. Однако, поскольку не проводилось прямое измерение персуазивности или ожиданий аудитории, интерпретации на основе данной теории носят предварительный характер и требуют дальнейшей эмпирической проверки.

Keywords: vaccine hesitancy, social media, digital communication, computational sociology, sentiment analysis, topic modeling

Ключевые слова: вакцинная нерешительность, социальные сети, цифровая коммуникация, вычислительная социология, анализ тональности, тематическое моделирование

Introduction

The resurgence of vaccine-preventable diseases represents a growing public concern, attracting considerable attention from the academic community. Central to this issue is vaccine hesitancy, a complex phenomenon driven by diverse beliefs that can lead to vaccine refusal [MacDonald, 2015]. Recognizing its impact, the World Health Organization has designated vaccine hesitancy as one of the top ten threats to global health, underscoring the need for a deeper investigation of immunization discourses [1].

Addressing this challenge requires a detailed understanding of the factors driving polarization in public attitudes. Existing research points to a complex interplay of determinants, including individual belief systems [Đorđević et al., 2021] and socioeconomic factors such as educational attainment and financial hardship [Anello et al., 2017]. At a broader level, the pervasive influence of the Internet and social media has been identified as a key catalyst, as digital platforms facilitate the rapid spread of vaccine-related misinformation, which substantially shapes public attitudes [Kata, 2010]. While scholars have increasingly examined anti-vaccine discourses on global platforms like Facebook* [2] and Twitter (X), a notable gap remains concerning Russian social media. In particular, there is a lack of theory-informed comparative analysis of the thematic and emotional content propagated by both anti-vaccination and pro-vaccination communities.

This study aims to address this gap by conducting a quantitative analysis of the digital discourse within the aforementioned communities on VK, a leading Russian social media platform. Building on evidence that the persuasive power of online movements depends on specific stylistic and thematic choices, we situate our investigation within the framework of Language Expectancy Theory (LET), employing it as an interpretive lens to generate hypotheses about expected linguistic and thematic differences between the two community types. This approach aligns with prior computational studies that have used LET to interpret observed language patterns without directly measuring audience perceptions. We posit that pro- and anti-vaccination communities are likely to exhibit distinct linguistic profiles tailored to the presumed expectations of their respective audiences. Thus, our research is designed to answer the following questions:

[1] Ten health issues WHO will tackle this year (2019) *World Health Organization*. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed: 22.09.2025].

[2] Readers should note that Meta Platforms, Inc. and its product Facebook* are legally designated as extremist organizations and banned in the Russian Federation. This context applies to all subsequent mentions in this text.

DOI: 10.14515/monitoring.2026.4.3147

— What are the **emotive** differences between pro- and anti-vaccination discourse on VK?

— What are the **thematic** differences between pro- and anti-vaccination discourse on VK?

By investigating the semantics and general sentiment of this content, our study provides a country-specific comparison of the online vaccination debate. We anticipate that this comparative understanding will contribute to a more robust sociological framework of vaccine hesitancy and acceptance, while simultaneously providing an impetus for further theoretically substantiated research in this domain. Ultimately, the findings could equip public health officials with the insights needed to develop targeted communication strategies that effectively promote vaccination and counter misinformation.

Related Work

Research Rationale

To situate this study within the current academic landscape, we turn to the sociology of media and communication. This field conceptualizes information as a contextual phenomenon that enables social actors to sustain complex networks and mutually influence one another through communicative processes. A key subfield, health communication, specifically investigates the media's role in framing health issues. Research in this area consistently demonstrates that the portrayal of health-related themes can create informational ambiguity (i. e. a lack of clarity or the presence of contradictory health messages), significantly impacting public awareness, behavior, and well-being. For instance, Renaud and colleagues argue that official information from public health entities is appropriated through engagement with media systems, which can influence health behaviors both positively and negatively [Renaud et al., 2006]. Other studies further demonstrate that media campaigns can shift individual perceptions of risk-related practices and alter group norms [Cohen, Scribner, Farley, 2000; Apollonio, Malone, 2009].

Vaccination is uniquely susceptible to media influence due to its dual nature as a scientifically complex issue and a contentious public debate [Larson et al., 2014]. This combination creates a fertile ground for misinformation, positioning immunization as a prominent and often polarized subject across media platforms. The dynamics of this relationship are starkly illustrated in Seth Mnookin's seminal work, «The Panic Virus», which analyzes the enduring but spurious link between vaccines and autism [Mnookin, 2011]. Mnookin argues that the media's institutional imperative for compelling narratives often reduces complex scientific evidence to the level of anecdote, thereby privileging emotionally charged stories over population-level research data [ibid.]. This tendency toward «false balance» [3], combined with the accelerated pace of digital information dissemination, has allowed the vaccine-autism myth to become deeply entrenched in public consciousness. Empirical research corroborates this analysis, demonstrating that traditional media coverage does not merely perpetuate misconceptions but can have a tangible impact on vaccine uptake [Recio-Román, Recio-Menéndez, Román-González, 2023].

[3] False balance is a media practice of presenting two opposing viewpoints as equally credible when one is supported by overwhelming scientific consensus and the other by fringe or discredited evidence.

DOI: 10.14515/monitoring.2026.4.3147

While these studies highlight traditional media, the Internet has emerged as the most prominent domain for influencing vaccination behavior. The Web 2.0 paradigm enables users to construct personal spaces within content-based networks [Murugesan, 2007], where individual experiences and frustrations become subjects of public discourse — a phenomenon particularly consequential in vaccine discussions [Kata, 2012]. Scholars theorize that the effective propagation of vaccine misinformation online is facilitated by easy access to information and a lack of consistent moderation by expert communities. As Betsch and colleagues note: «by actively creating and disseminating information, users become more involved, which is assumed to amplify potential effects of information on perceptions, attitudes, and behavior» [Betsch et al., 2012: 3728]. This active participation adds a cognitive dimension to understanding how social media influences vaccine decision-making, potentially leading individuals to believe that the risks of vaccine side effects outweigh the risks of vaccine-preventable diseases. Psychosociological research further suggests that influential social media actors (e. g. influencers) can impose their interpretations of health concerns on various groups, shaping their attitudes [Powell, Pring, 2024]. This is especially evident when individuals who have had negative vaccination experiences seek information to confirm their pre-existing beliefs [Malthouse, 2023].

Given that social media facilitates the sharing of personal vaccination narratives, a careful analysis of this content is essential. However, while a substantial body of research has examined vaccine discourse on English-language platforms like Facebook* [Faasse, Chatman, Martin, 2016], Twitter (X) [Broniatowski et al., 2018], and YouTube [Tokojima Machado, de Siqueira, Gitahy, 2020], Russian-specific social media remains relatively underexplored. The existing studies on Russian networks often employ a technical, rather than descriptive, approach to content analysis. For example, Tserkovnaya and Larson used a classification algorithm to find that 59 % of vaccine-related posts on VK express anti-vaccination sentiment, with child health as a dominant topic [Tserkovnaya, Larson, 2019]. In contrast, a recent thematic analysis of user comments on VK identified the core structure of vaccine hesitancy discourse as being organized around four key themes: the riskiness of vaccination, its ineffectiveness, its unnecessary nature, and its perceived unfairness [Dudina, Saifulina, 2023]. Vaccination has also been addressed within studies on conspiracy ideologies, with evidence suggesting that COVID-19 immunization is framed by vaccine hesitant users as a planned «microchipping» operation [Platonov, Svetlov, 2020]. Despite these valuable contributions, a comparative analysis of pro- and anti-vaccination discourse on VK informed by Language Expectancy Theory has not been conducted, leaving a significant gap in the literature.

Language Expectancy Theory (LET)

Language Expectancy Theory (LET) provides a sociological framework for understanding persuasion, positing that language is a rule-governed system shaped by cultural and social norms [Burgoon, 1995]. Through socialization, individuals develop shared expectations, regarding appropriate linguistic behavior for specific situations and speakers [Averbeck, Miller, 2014]. According to LET, persuasive success depends on a message's strategic alignment with or deviation from these normative

DOI: 10.14515/monitoring.2026.4.3147

expectations, thereby influencing audience attitudes and behaviors [Burgoon, Pauls, Roberts, 2002].

LET has been applied across a range of communicative contexts beyond its original face-to-face settings. For instance, Jensen and colleagues extended LET to the online environment, demonstrating that textual properties of anonymous product reviews produce expectancy violations influencing credibility attributions [Jensen et al., 2013]. They showed that even without traditional source cues, language characteristics alone can trigger expectancy violations, establishing a precedent for applying LET to technology-mediated communication. Lee and Yu further advanced this application by using LET to analyze how linguistic style and content features of disaster tweets predict resharing behavior on Twitter [Lee, Yu, 2020], operationalizing LET constructs computationally through natural language processing tools. Similarly, Liu, Jiang, and Wang applied LET to online health consultations, examining how physicians' writing style affects patient satisfaction [Liu, Jiang, Wang, 2023].

These studies demonstrate that LET can be applied to digital communication contexts. However, whereas Jensen et al. [2013] directly measured audience expectations through a controlled experiment, Lee and Yu [2020] and Liu, Jiang, and Wang [2023] inferred expected norms from observed language patterns. Our study follows the latter approach: we use LET as an interpretive framework to generate theoretically grounded hypotheses, meaning our claims about alignment with audience expectations remain interpretive rather than empirically validated at the audience level.

Emotional intensity of the language

One key dimension highlighted by LET is the use of emotionally intense language. In health communication, such intensity can serve as a powerful persuasive tool by making abstract risks feel immediate and tangible [Vranceanu et al., 2012]. For instance, Buller and colleagues demonstrated that vivid descriptions of skin melanoma risks effectively encouraged preventive behaviors among parents [Buller et al., 2000].

We posit that the socio-psychological profiles of pro- and anti-vaccination communities are likely to be associated with divergent communicative patterns, which are reflected in their use of emotional language. Research indicates that anti-vaccination discourse is often driven by high-risk perception [Reyna, 2012] and conspiracy ideation [Jolley, Douglas, 2014], which thrive on a sense of urgency and outrage. Given these characteristics, we expect anti-vaccination content to exhibit more intense, negative language that would be consistent with the emotional orientation of its audience. In contrast, pro-vaccination discourse, which tends to align with the authoritative, evidence-based discourse of public health, is more likely to adhere to norms of rationality and analytical detachment [Faasse, Chatman, Martin, 2016]. From a LET perspective, this would represent conformity to the expected norms of institutional and scientific communication. This leads to our first two hypotheses:

H1: Anti-vaccination posts will demonstrate higher emotional intensity (a greater frequency of sentiment-laden words) compared to pro-vaccination posts.

H2: Anti-vaccination posts will be characterized by a more negative emotional tone compared to pro-vaccination posts.

DOI: 10.14515/monitoring.2026.4.3147

Thematic variability

Within the LET framework, audience expectations extend beyond emotional tone to encompass the substantive content of communication. The topics a community prioritizes can be understood as reflecting the contextual and semantic expectations of its audience — that is, members come to expect certain thematic orientations as normative for their community. From this perspective, thematic analysis serves as a complementary dimension of LET: just as emotional intensity must align with audience expectations to be effective, so too must the topical focus of discourse resonate with the core values and concerns of its intended audience. Comparative studies have consistently demonstrated thematic disparities between pro- and anti-vaccination communities. For example, pro-vaccination content often frames immunization as a collective social responsibility, whereas anti-vaccination discourse frequently emphasizes individual bodily autonomy and alleged harms [Faasse, Chatman, Martin, 2016].

In the Russian context, we anticipate that these global trends will be inflected by local specificities. Petrov's finding that anti-vaccination content on VK frequently focuses on children and family suggests a thematic orientation leveraging culturally salient values — a potential difference from Western patterns that underscores the need for context-sensitive analysis [Petrov, 2022]. Furthermore, aligning with global patterns, we expect Russian anti-vaccination discourse to employ conspiracy theories as a frame for interpreting events [Platonov, Svetlov, 2020]. Conversely, pro-vaccination communities are expected to leverage topics related to science and public health efficacy to build credibility and conform to expectations of an institutional voice. While we formulate these expectations as separate hypotheses for analytical clarity, we acknowledge that the vaccination debate in Russia has activated a broad range of adjacent themes reflecting wider social tensions. The hypotheses below capture the most theoretically grounded expectations rather than an exhaustive list of all possible thematic differences:

H3: Topics in anti-vaccination communities will be more prominently related to family and children than topics in pro-vaccination communities.

H4: Topics in anti-vaccination communities will be more prominently related to conspiracy theories than topics in pro-vaccination communities.

H5: Topics in pro-vaccination communities will be more prominently related to science and education than topics in anti-vaccination communities.

Data

Community identification and sampling

This study conducts a comparative analysis of two distinct sets of VK communities: one advocating for and one opposing vaccination. The communities in our sample are predominantly grassroots, self-organized initiatives created and maintained by ordinary users, though the sample also includes some official pages affiliated with medical or public health institutions.

Anti-vaccination communities: the initial sample was drawn from a list of communities that VK had officially flagged with a specialized warning about potentially unreliable information. When users attempted to access these pages, a warning message [4]

[4] As of 2026, the warning message will no longer appear. The VK has decided to stop this practice without providing an official explanation.

DOI: 10.14515/monitoring.2026.4.3147

was displayed, accompanied by a link to a World Health Organization article on vaccination [Petrov, 2022]. In 2022, colleagues from VK provided us with the complete list of these flagged communities, which served as the foundation for our anti-vaccination sample. The original dataset comprised 282 public pages. From this list, 46 restricted communities were excluded due to inaccessible content. A further 7 communities were removed because they were either blocked by the platform or contained no vaccination-related material, resulting in a final sample of 229 anti-vaccination pages. Examples of anti-vaccination communities in our sample include: «Мамы и папы против прививок!!» («Moms and Dads Against Vaccinations!!»), «Мой ребенок без прививок» («My Child Is Unvaccinated»), «Вакцинация — панацея или жертвоприношение?» («Vaccination: Panacea or Human Sacrifice?»).

Pro-vaccination communities: a comparable set of pro-vaccination communities was constructed using a subscriber intersection algorithm via vk.barkov.net. The essence of the algorithm is that, starting from several pro-vaccination communities, it identifies thematically similar pages where subscribers from the original communities tend to appear. This strategy resembles a snowball sampling technique. Three prominent pro-vaccine communities were used as seeds to generate three separate lists of the top 100 related groups. After combining these lists and removing duplicates, the resulting communities were screened for relevance. Using a list of 10 vaccination-related keywords, a community was retained only if at least one keyword appeared in its posts. This process yielded a final sample of 84 pro-vaccination pages. Examples of pro-vaccination communities in our sample include: «О прививках без истерик» («About Vaccines Without Hysteria»), «Прививка от мракобесия 18+» («Vaccine Against Obscurantism 18+»), «Серьёзно о прививках и не только» («Seriously About Vaccinations and Beyond»).

It should be noted that by design, all posts collected from anti-vaccination communities are classified as anti-vaccination content, and all posts from pro-vaccination communities are classified as pro-vaccination content. While individual posts within a community may occasionally express dissenting views, the community-level classification reflects the dominant orientation of each page, as established by VK's labeling system (for anti-vaccination communities) and the thematic screening procedure (for pro-vaccination communities).

Data filtering and pre-processing

To ensure semantic relevance of the data, posts from both community sets were subjected to a multi-stage filtering and cleaning process, detailed below and summarized in table 1.

Keyword filtering, deduplication, and text cleaning were applied sequentially to both datasets. First, as the pro-vaccination groups often addressed general medical topics, a uniform keyword filter was used to isolate posts specifically about vaccination: a post was retained for analysis only if it contained at least one of 30 vaccination-related keywords. Next, duplicate posts within each dataset were identified and removed, and all posts containing fewer than 50 words (including those with no text) were excluded to minimize noise and significant disparities in post length. Finally, the text of the remaining posts was lemmatized and cleaned of non-essential elements, including English words, emoticons, and a standard list of 568 Russian stop words.

DOI: 10.14515/monitoring.2026.4.3147

Final dataset construction

The initial raw data and the results of each filtering step for both community types are presented in table 1. After the respective filtering pipelines were completed, the pro- and anti-vaccination datasets were merged, and an ID variable was added. The combined dataset initially contained 18,922 posts. A final check for duplicates across the entire corpus identified and removed 683 redundant texts. The final analytical dataset consists of 18,239 unique posts described by 17 variables. This total comprises 10,987 posts from anti-vaccination communities and 7,252 posts from pro-vaccination communities. All posts were collected via the VK API, which returned the 1,000 most recent posts per community. As a result, the sample is skewed toward more recently published content. The posts in the final dataset span the period from November 2007 to April 2022.

Table 1. ***Data Filtering Pipeline and Sample Sizes***

Processing step	Anti-vaccination dataset	Pro-vaccination dataset
Initial Raw Posts	90,446	67,314
After Deduplication	44,352	53,990
After Length Filtering (>50 words)	22,147	25,257
After Keyword Filtering (Final)	11,204	7,718
After Merging & Cross-Deduplication	10,987	7,252
Total	18,239	

Methodology

To investigate the thematic and emotive differences between pro- and anti-vaccination discourse on VK, this study employs a computational approach, combining structural topic modeling (STM) with two dictionary-based sentiment analysis techniques.

Sentiment analysis

To address the first research question concerning emotional differences, we utilized two dictionary-based approaches to generate sentiment ratings for the textual data.

First, the Linis-crowd.org dictionary [Koltsova, Alexeeva, Kolcov, 2016], a pre-existing Russian sentiment lexicon, was employed to generate polarized sentiment scores for each post. This dictionary comprises 7,545 words, each assigned a score ranging from −1 (negative) to 1 (positive). The analysis procedure involved tokenizing the pre-processed texts and joining them with the lexicon. The initial tokenized dataset contained 4,045,236 observations. Following an inner join by word, this was reduced to 1,692,355 matched tokens. The number of unique word matches was 6,987 out of 7,545, indicating a 92.6 percent congruence with the dictionary. The overall emotional tone of each post was defined as the average sentiment rating of all the matched words within the text. To test for significant mean differences in sentiment between the pro- and anti-vaccination groups, we employed the Welch Two Sample t-test and the Wilcoxon rank sum test and calculated the respective effect sizes.

DOI: 10.14515/monitoring.2026.4.3147

Second, to move beyond simple polarity and capture a broader spectrum of affective states, we used the Russian version of the NRC Emotion Lexicon [Mohammad, Turney, 2013]. This lexicon categorizes words into eight basic emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The authors cautioned that the Russian version was created via automated translation. Following manual assessment, we removed 477 untranslated entries, 72 transliterations, and 759 duplicates, proceeding with a refined lexicon of 8,613 words. We then tokenized the post texts and merged them with the curated NRC dictionary. The tokenized database of 4,045,236 observations was reduced to 1,751,518 after merging. The number of unique word matches was 3,546 out of 8,613, showing a 41.17 percent congruence. To identify the predominant emotions in pro- and anti-vaccination posts, we assessed the frequency of terms associated with each of the eight emotions. The Welch Two Sample t-test and the Wilcoxon rank sum test were again used to test for significant differences in the prevalence of each emotion between the two groups.

Structural topic modeling

To assess the thematic variability within the pro- and anti-vaccination communities and address our second research question, we employed Structural Topic Modeling (STM), an unsupervised machine learning technique. STM was selected over conventional Latent Dirichlet Allocation (LDA) because it can model theme prevalence as a function of document-level metadata [Blei, Ng, Jordan, 2003]. This capability is particularly relevant for social media data, where text production is influenced by various external factors, such as community affiliation.

The STM procedure was implemented in several stages. First, the cleaned and lemmatized data was processed using the *textProcessor* function, which removed 53,317 of the 108,710 unique terms based on low frequency. Second, documents, vocabulary, and metadata were prepared using the *prepDocuments* function to correct for zero-indexing and remove unused words. The resulting processed corpus contained 18,239 documents, 55,393 unique terms, and 2,709,763 tokens. Third, to identify the optimal number of topics, a set of four models was estimated with topic counts (K) of 10, 20, 30, and 40. The model with 40 topics was selected based on diagnostic statistics, including held-out likelihood and residuals [Taddy, 2012]. Finally, the prevalence of each topic was assessed using expected topic proportions. The distribution of topics across the pro- and anti-vaccination communities was determined using the *estimateEffect* function, which performs a series of regression models with topic proportion as the dependent variable and group metadata as the predictor.

For the purposes of academic transparency, we also wish to point out that the identification and labeling of topics in STM involves a standard two-stage process. First, the algorithm statistically extracts topics based on patterns of word co-occurrence across documents. Second, researchers interpret these statistical outputs (examining lists of high-frequency words for each topic and reading a sample of representative posts using the *findThoughts* function) to assign meaningful labels. This second stage is inherently interpretive and involves researcher judgment. However, this combination of computational extraction and qualitative interpretation is a widely accepted practice in academic literature [Weston et al., 2023]. In the results section, we present both

DOI: 10.14515/monitoring.2026.4.3147

the statistical topic prevalences and our interpretive labels, acknowledging that different researchers might assign slightly different labels to the same word clusters.

Results and Discussion

Quantifying sentiment polarization

First, we tested for a statistically significant difference in mean sentiment between pro- and anti-vaccination groups. We employed a t-test with Welch's adjustment to account for the unequal variances between the groups. Simultaneously, acknowledging the non-normal distribution of the sentiment data, we verified the results using a non-parametric alternative. The Welch Two Sample t-test indicates a statistically significant, negative, and very small effect (*difference = –0.02, 95 % CI [–0.02, –0.01], t = –7.67, p < .001; Cohen's d = –0.12, 95 % CI [–0.15, –0.09]*). The non-parametric test supports this conclusion (*W = 36663622, p < .001; r (rank biserial) = –0.08, 95 % CI [–0.10, –0.06]*). Given the very small effect sizes for both Cohen's d and the rank-biserial correlation, the identified association should be treated with caution. The key takeaway is that the sentiment of vaccination-related material on VK does differ between supporters and opponents, with vaccine opponents using slightly more negative language. It is also worth noting that although the sentiment is negative, the emotionality tends to hover around zero, which might be due to the length of the input texts. Viewed through a LET lens, this pattern is consistent with the expectation that anti-vaccination communities, oriented toward audiences with heightened risk perception, would employ more negatively charged language, while pro-vaccination communities would adhere to the more measured register associated with institutional and scientific communication. However, since we did not directly measure audience expectations, this interpretation remains tentative.

Contrasts in negative language

The top terms associated with negative sentiment were analyzed to further assess the observed difference. We identified the most frequent terms with a sentiment score less than –0.3 to illustrate the characteristic negative vocabulary (see table 2).

The results show that the negative language used by those who oppose vaccination is closely associated with the potential side effects of immunization, as seen in terms like *болезнь / заболевание* (illness) and *осложнение* (complication). Crucially, the concordance[5] data reveal that these terms are typically embedded in narratives of personal harm and systemic concealment. For instance, one anti-vaccine post states: *«замалчивание поствакцинальных осложнений и откровенное преувеличение опасности детских болезней!» («the concealment of post-vaccination complications and the blatant exaggeration of the danger of childhood illnesses!»)*, while another reads: *«у здоровых детей прививка проявляется с серьёзными осложнениями, это здоровые детки становятся инвалидами» («in healthy children the vaccination manifests with serious complications, healthy children become disabled»)*. These examples illustrate how negative medical vocabulary in anti-vaccine discourse serves to construct vaccination itself as the source of disease, rather than its remedy.

[5] In text mining, a concordance is a list that shows every time a particular word or phrase appears in a document (or collection of documents), along with the words that come before and after it.

DOI: 10.14515/monitoring.2026.4.3147

Furthermore, this group demonstrates a propensity for harsh language, employing expressions such as *умирать* (to die), *убивать* (to kill), and *смерть* (death). The concordances confirm that these terms are used to attribute lethal consequences directly to vaccines. A representative post warns: *«Прививка может сделать Вашего ребёнка калекой или убить» («A vaccination can make your child disabled or kill them»)*. Such language constructs an existential threat frame around vaccination, positioning it as a life-or-death matter. In contrast, proponents of vaccination use negative language that primarily refers to the consequences of vaccine hesitancy, as indicated by terms like *пневмония* (pneumonia), *тяжелый* (severe), and *риск* (risk). The concordance analysis shows that these words appear in clinical and epidemiological contexts. For example, one pro-vaccine post explains: *«Другого способа уберечь вас от тяжёлой болезни просто не существует» («There is simply no other way to protect you from severe illness»)*, while another notes: *«одно из вероятных осложнений — пневмония, которая может стоить жизни» («one of the probable complications is pneumonia, which can cost a life»)*. Here, negative vocabulary is directed outward toward the diseases that vaccines prevent rather than toward the vaccines themselves.

It is also worth noting that supporters frequently use words like *боязнь* (phobia) and *страх* (fear), which connect to the group's focus on explaining the reasons behind vaccine refusal. The concordances show that these terms are used analytically rather than emotionally: *«Незнание, боязнь побочных эффектов, как правило, снижают охват вакцинацией» («Ignorance, fear of side effects, as a rule, reduce vaccination coverage»)*, and *«В последние годы заболеваемость корью выросла из-за боязни прививок» («In recent years, measles incidence has risen due to phobia of vaccinations»)*. In both cases, fear is presented as a barrier to be overcome through education, not as a justified reaction.

Table 2. ***Top-10 Negative Sentiment Words in Pro- and Anti-vaccination Content***

Anti-vaccination	Pro-vaccination
Vaccination I Вакцинация (*n* = 599)	Infection I Инфекция (*n* = 850)
To die I Умирать (*n* = 461)	Illness I Болезнь (*n* = 650)
Illness I Болезнь (*n* = 408)	Illness I Заболевание (*n* = 635)
Infection I Инфекция (*n* = 385)	Fear I Боязнь (*n* = 293)
Death I Смерть (*n* = 350)	Therapy I Терапия (*n* = 250)
Illness I Заболевание (*n* = 343)	Treatment I Лечение (*n* = 240)
To kill I Убивать (*n* = 219)	Vaccination I Вакцинация (*n* = 202)
Virus I Вирус (*n* = 207)	Pneumonia I Пневмония (*n* = 181)
Complication I Осложнение (*n* = 201)	Risk I Риск (*n* = 181)
To become ill I Заболевать (*n* = 200)	Severe I Тяжелый (*n* = 179)

Predicting users' engagement with sentiment

Second, simple linear regression was employed to assess whether the overall emotional tone of the text is associated with increased post engagement. The dependent variable was operationalized as the total number of likes, comments, and reposts, while the predictors included content type (pro-vaccination or anti-vaccination) and sentiment (ranging from −1 to 1). Before fitting the model, cases identified as outliers based on the dependent variable were removed using the interquartile range method, resulting in the exclusion of 2,114 observations. Consequently, the regression model was fitted using 16,125 observations: 9,942 from the anti-vaccination group and 6,183 from the pro-vaccination group. All continuous variables were z-score standardized, and variance inflation factors indicated no multicollinearity concerns (all VIF < 1.80).

The model explains a statistically significant and very weak proportion of variance ($R^2 = 0.01$, *F (3,16121) = 70.59, p < .001, adj.* $R^2 = 0.01$). The model's intercept, corresponding to sentiment = 0 and type = anti-vaccination, is at −0.09 (*95 % CI [−0.11, −0.07], t (16121) = −8.83, p < .001*). Within the model:

The effect of sentiment is statistically non-significant and negligible in magnitude (*beta = 0.003, 95 % CI [−0.02, 0.02], t (16121) = 0.36, p = 0.718*), indicating that emotionality of the post does not predict engagement in the given sample.

The effect of type [pro-vaccination] is statistically significant and positive (*beta = 0.23, 95 % CI [0.20, 0.26], t (16121) = 14.14, p < .001*), indicating that, holding sentiment constant at its mean, pro-vaccination posts have a standardized engagement score that is, on average, 0.23 standard deviations higher than that of anti-vaccination posts.

The interaction effect of type [pro-vaccination] on sentiment is statistically non-significant and positive (*beta = 0.03, 95 % CI [−0.00259, 0.06], t (16121) = 1.80, p = 0.073*), meaning that an association between post engagement and sentiment is not moderated by type of the content.

Overall, the model explains a relatively small portion of the variance ($R^2 = 0.01$) and, as such, has limited explanatory power. Therefore, while the analysis suggests that sentiment alone does not predict engagement and that pro-vaccination posts are associated with higher engagement, these findings should be interpreted with caution. The substantial amount of unexplained variance indicates that other factors not included in this model likely play a major role in driving user engagement with posts. Additionally, the aggregation of likes, comments, and reposts into a single composite engagement score may obscure platform- and metric-specific dynamics. Future research would benefit from analyzing these indicators separately.

An exploratory analysis of emotions with the NRC Lexicon

As a second dictionary-based method, we used the Russian version of the NRC Emotion Lexicon. This approach involved examining the frequency of phrases associated with eight distinct emotions to determine whether there were any statistically significant differences between the anti- and pro-vaccine groups. Prior to statistical testing, we visually analyzed the predominance of emotions and determined that fear, trust, and anger were the most prominent within each group. Surprisingly, no noticeable differences in overall emotion prevalence were found — both the anti-vaccine and pro-vaccine groups used emotion-related words in similar proportions across their content.

DOI: 10.14515/monitoring.2026.4.3147

A closer examination of the concordance data allowed us to characterize the specific terminology linked to these emotions and, crucially, to understand how these terms functioned within the vaccination debate. First, both anti-vaccine and pro-vaccine messaging included anger-related terms referencing positions of responsibility. However, opponents of vaccination were more likely to employ phrases indicating personal agency, such as *отказ* (refusal), *отказывать* (to refuse), and *защита* (protection). In the concordance data, these terms appeared in contexts where individuals asserted their right to reject vaccination — for example, *«право на отказ от вакцинации» («the right to refuse vaccination»)* and *«защита прав здоровых детей и их родителей» («protection of the rights of healthy children and their parents»).* In contrast, proponents used phrases like *нарушение* (violation), *решать* (to decide), and *борьба* (fight), which in context referred to collective action against disease — e. g., *«человечество против пневмококка — история борьбы и побед» («humanity vs. pneumococcus: the history of the fight and victory»)*, *«нарушение графика вакцинации» («violation of the vaccination schedule»)*, and *«способ борьбы с ОРВИ и гриппом» («a method of fighting acute respiratory infections and influenza»)*. This contextual distinction suggests that anti-vaccine vocabulary reflects a preference for individual-level autonomy (the freedom to make personal medical decisions), whereas pro-vaccine rhetoric frames vaccination as a matter of collective duty, where deviations from recommended schedules or refusals constitute violations of shared responsibility for public health. From a LET perspective, these divergent framings are consistent with the idea that each community type employs anger-related vocabulary in ways that align with the core values presumed to be normative for its audience: personal freedom for anti-vaccination communities and collective responsibility for pro-vaccination ones.

Second, we found that trust-related terms also co-occurred with different thematic vocabulary in each group, pointing to distinct discursive contexts in which trust is discussed. Among opponents, trust-related vocabulary frequently co-occurred with references to the state and legal system, as seen in words like *власть* (power/authority) and *закон* (law). Concordance analysis shows that these terms appeared in contexts where institutional authority and legal frameworks were being questioned or contested — for instance, *«народ не доверяет системе власти» («the people do not trust the system of power»), «по закону никто не может обязать вакцинацию» («by law, no one can mandate vaccination»)*. This does not allow us to conclude that opponents distrust the state per se, but it does indicate that their discourse about trust is predominantly situated within an institutional and legal register. The vocabulary of trust, in these contexts, serves as a rhetorical resource for contesting authority rather than expressing confidence in it. In contrast, among vaccination supporters, trust-related vocabulary co-occurred with terms associated with the educational system — *школа* (school) — and the maternal figure — *мама* (mother). Concordance examples reveal that these terms appeared in contexts of everyday parenting decisions: discussions about whether schools require vaccination certificates, mothers' concerns about their children's health, and references to pediatric recommendations. Within a LET framework, this divergence in the discursive environments surrounding trust-related vocabulary is noteworthy. It suggests that the two communities construct the relevance of trust through different rhetorical strategies: anti-vaccination discourse situates trust within a macro-institutional frame potentially resonating with an audience attuned to questions of state power and individual rights,

DOI: 10.14515/monitoring.2026.4.3147

while pro-vaccination discourse situates trust within a micro-relational frame of family and schooling, potentially resonating with an audience oriented toward routine health-seeking practices. Importantly, these are observations about how trust is discursively framed in each community, not about the actual trust dispositions of their members.

Finally, we determined that phrases related to fear commonly reflected scenarios of sickness and admission to medical institutions, evidenced by words such as *больной* (ill/patient), *болезнь* (illness), and *больница* (hospital). Notably, both groups shared this fear-related vocabulary, but the concordances reveal divergent framings. In anti-vaccine texts, illness-related terms appeared alongside language minimizing disease severity — e. g., *«корь, краснуха — это не такие страшные болезни, не бубонная чума, чтобы рисковать» («measles, rubella — these are not such terrible diseases, not the bubonic plague, to take risks [with vaccination]»)*. In pro-vaccine texts, by contrast, the same vocabulary was used to emphasize the genuine threat of disease — e. g., *«другого способа уберечь вас от тяжёлой болезни просто не существует» («there is simply no other way to protect you from severe illness»)* and discussions of hospitalization and complications. Thus, while fear-related language is shared across both camps, it serves opposite rhetorical functions: for opponents, it is deployed to argue that the diseases themselves are not fearsome enough to justify vaccination risks, while for proponents, it underscores the very real danger that vaccination is designed to prevent. This finding highlights a key insight for LET-informed analysis: identical emotional categories can carry divergent communicative orientations depending on the community.

Statistical analysis of emotions with the NRC Lexicon

Finally, eight t-tests were performed to determine if there was a statistically significant difference in the frequency of emotion-related phrases between the pro- and anti-vaccine groups. We used the Welch Two Sample t-test, as was done with the Linis-crowd.org sentiment analysis, and verified the results using the Wilcoxon rank sum test. Only the parametric test results will be presented in this article (see table 3). Non-parametric test results are available in the Supplementary Materials.

Table 3. ***Results of The Welch Two Sample t-test***

Emotion	Mean (antivax)	Mean (provax)	Statistic	p-value	CI	Results	Effect size
anger	12.33	7.55	24.53	<.001	[4.40, 5.16]	significant	0.35 [small]
anticipation	13.84	10.74	14.17	<.001	[2.67, 3.53]	significant	0.21 [small]
disgust	8.67	6.58	13.17	<.001	[1.78, 2.40]	significant	0.19 [very small]
fear	17.88	14.24	12.23	<.001	[3.05, 4.22]	significant	0.18 [very small]
joy	10.37	10.11	1.26	0.206	[−0.14, 0.66]	non-significant	0.02 [very small]
sadness	12.18	10.02	9.87	<.001	[1.73, 2.59]	significant	0.15 [very small]
surprise	5.39	4.07	13.17	<.001	[1.12, 1.52]	significant	0.19 [very small]
trust	25.50	17.39	21.70	<.001	[7.38, 8.84]	significant	0.31 [small]

DOI: 10.14515/monitoring.2026.4.3147

The results, based on a significance level of 0.05, indicate that the frequency of emotion-specific words differs between opponents and proponents of vaccination for all emotions except joy (p-value = 0.206). Specifically, anti-vaccine content is characterized by a greater frequency of phrases related to sentiments identified in the NRC dictionary. Based on effect size, this difference is particularly noticeable for anger (Cohen's d = 0.35), trust (Cohen's d = 0.31), and anticipation (Cohen's d = 0.21). Despite the statistical significance of these results, the effect size in both parametric and non-parametric tests remains small. The key conclusion is that anti-vaccination content contains more emotion-related vocabulary.

Uncovering key topics with STM

The analysis of expected topic proportions, conducted on a dataset combining pro- and anti-vaccine posts, identified five predominant themes: COVID-related restrictions (Topic 14), Sources of Information (Topic 12), the Right to Refuse Vaccines (Topic 18), the COVID Vaccine Itself (Topic 2), and issues concerning Children and Illness (Topic 29). Visualizations of the top five topics by expected topic proportion, along with visualizations of the other topics identified in the analysis, are available in the Supplementary Materials. These prominent topics reveal several key characteristics of the discourse. First, the pandemic itself appears to have shaped a distinct debate focused on governmental policies, a trend clearly embodied by Topic 14, which encompasses discussions of vaccination, mask mandates, and lockdowns. Second, the significance of Topic 12 (Sources of Information) highlights a strong tendency among participants to actively produce and disseminate content, such as articles, videos, and posts, rather than just consume it. Third, the prominence of Topic 29 (Children and Illness) underscores how children's health served as a central focal point, often used to debate broader medical concerns, including the effects of vaccination. Finally, the high prevalence of Topic 18 (Right to Refuse Vaccines) indicates that legal frameworks and exemptions allowing vaccine refusal held a particularly influential position within the overall conversation. Overall, these findings provide a preliminary map of the main thematic axes within the overall discussion, while any differential emphasis between subgroups remains to be examined in the subsequent prevalence analysis.

Analyzing topic prevalence across groups with STM

To identify topics that were more prevalent in one group than the other, we built a series of regression models. These models used topic proportion as the dependent variable and content type as a predictor, with the anti-vaccination group serving as the reference. The visualization of the results is available in the Supplementary Materials.

The model summary indicates that only 35 topics exhibited a statistically significant difference between the groups at the 0.01 significance level. Several topics showed insignificant results, including Topic 17 (Immune System), Topic 31 (Big Pharma), Topic 38 (Educational System), and Topic 40 (Role of Vaccination). This suggests that these ideas are discussed with similar frequency by both groups.

The analysis also revealed several notable differences. Most coronavirus-related topics, including Topic 2 (COVID Vaccine), Topic 5 (Anti-COVID Movement), Topic 14

DOI: 10.14515/monitoring.2026.4.3147

(COVID-related Limitations), and Topic 19 (Pandemic), were concentrated in the anti-vaccination content. The exception was Topic 16 (Research on COVID), which appeared more frequently in pro-vaccination publications.

Similarly, topics concerning adverse effects of vaccination, such as Topic 6 (General Vaccination), Topic 11 (Mortality and Vaccines), and Topic 26 (Vaccine as Source of Illness), were predominant in the anti-vaccine group. In contrast, the pro-vaccine group more frequently discussed topics related to infections and their symptoms, such as Topic 8 (Infections and Treatment) and Topic 37 (Symptoms of Illnesses). Finally, topics like Topic 34 (External Control) and Topic 39 (Religious Figures), which encompass conspiracy theories and religion, appear to be of particular interest to vaccine opponents. These thematic contrasts support H3, H4, and H5 and are interpretable through LET: anti-vaccination communities concentrate on topics that might resonate with an audience oriented toward distrust of institutions and concern for family welfare, while pro-vaccination communities might prioritize topics aligned with the expectations of an audience that values scientific authority and evidence-based public health. The comparative framework thus reveals not only what each side discusses but also how these thematic choices may reflect distinct communicative orientations tailored to different audience norms. Although further research is needed to establish direct causal links between audience expectations and content production strategies.

Conclusion

This study mapped the emotive and thematic landscape of the vaccination debate on VK, using Language Expectancy Theory (LET) as an interpretive lens. Our computational analysis of 18,239 posts from 313 communities reveals systematic differences in both emotional tone and thematic focus between pro- and anti-vaccination communities.

Regarding emotional language, our findings confirm H1 and H2: anti-vaccination posts exhibit higher emotional intensity and a more negative tone, though effect sizes are small. Concordance analysis reveals that anti-vaccination content employs negative vocabulary such as *смерть* (death), *убивать* (to kill), and *осложнение* (complication) within texts framing vaccination as an existential threat, while pro-vaccination content directs negative vocabulary toward diseases vaccines prevent, using fear-related terms analytically. From a LET perspective, anti-vaccination discourse employs emotionally intense language consistent with an audience oriented toward high-risk perception, whereas pro-vaccination discourse adheres to the measured register of institutional communication. However, since audience expectations were not directly measured, these interpretations remain hypothetical.

Thematically, STM results support H3, H4, and H5. Anti-vaccination communities more prominently feature topics related to children and family, conspiracy theories, and religious figures, while pro-vaccination communities emphasize scientific research and disease prevention. The NRC emotion analysis further revealed that shared emotional categories (anger, trust, fear) serve divergent rhetorical functions: anti-vaccination anger vocabulary centers on individual autonomy, while pro-vaccination anger vocabulary frames vaccination as collective action against disease. Trust-related terms in anti-vaccination discourse mark institutional distrust, whereas in pro-vaccination discourse they are grounded in relational contexts such as family and school.

Our findings also suggest context-specific features of the Russian debate, including the prominence of legal frameworks for vaccine refusal and the centrality of children and family in anti-vaccination discourse, though direct cross-national comparisons would be needed to confirm these distinctions. A critical finding is that emotional tone does not predict user engagement, although post type (pro- vs. anti-vaccination) shows a small but significant association, with pro-vaccination posts receiving slightly higher engagement, suggesting that other factors such as community size, content format, or network structure may play a more substantial role in driving interaction.

Taken together, these findings document systematic differences in the linguistic profiles of pro- and anti-vaccination communities on VK. The comparative framework reveals that these differences are not limited to surface-level metrics but extend to how shared linguistic resources such as fear, trust, and anger vocabulary are deployed in contextually distinct ways. These patterns are consistent with LET-generated expectations: each community type exhibits a linguistic and thematic profile that aligns with the presumed normative orientation of its audience. However, we emphasize that our study identifies what differences exist and interprets them through a theoretical lens, rather than empirically demonstrating that these patterns function as persuasive strategies or that they are perceived as such by the audiences. Establishing causal and perceptual links, including direct measurement of audience expectations and attitude change, would require experimental or survey-based designs, which we identify as a priority for future research.

Limitations

Several limitations shape the scope and interpretation of our findings. First, while we employed LET as an interpretive lens following prior computational studies, core LET constructs (audience expectations, expectancy violations, and persuasive outcomes) were not empirically measured. Our claims about alignment with audience expectations are therefore interpretive inferences. Alternative explanations, such as differences in community demographics or content moderation, cannot be ruled out. Future research should employ experimental designs to directly assess whether the identified linguistic patterns influence attitudes or behavior.

Second, the VK API limited retrieval to 1,000 most recent posts per community, potentially introducing recency bias. The asymmetry in sample sizes (229 anti-vaccination vs. 84 pro-vaccination communities) reflects VK's structural landscape rather than a design choice but may affect generalizability. Our findings should be understood as a cross-sectional snapshot rather than a longitudinal account.

Third, the dictionary-based sentiment methods presented specific challenges. The Linis-crowd approach, which averages word-level scores across posts, is sensitive to text length and may attenuate emotional differences — likely contributing to the small observed effect sizes. The relatively low congruence rate of the NRC Emotion Lexicon with our corpus (41.17 %) means that emotion-specific findings should be interpreted with caution. The development of sentiment lexicons tailored to Russian-language health discourse would substantially improve future analyses.

DOI: 10.14515/monitoring.2026.4.3147

References (Список литературы)

1. Anello P., Cestari L., Baldovin T., Simonato L., Frasca G., Caranci N., Pascucci M. G., Valent F., Canova C. (2017) Socioeconomic Factors Influencing Childhood Vaccination in Two Northern Italian Regions. *Vaccine*. Vol. 35. No. 36. P. 4673—4680. https://doi.org/10.1016/j.vaccine.2017.07.058.

2. Apollonio D. E., Malone R. E. (2009) Turning Negative into Positive: Public Health Mass Media Campaigns and Negative Advertising. *Health Education Research*. Vol. 24. No. 3. P. 483—495. https://doi.org/10.1093/her/cyn046.

3. Averbeck J. M., Miller C. (2014) Expanding Language Expectancy Theory: The Suasory Effects of Lexical Complexity and Syntactic Complexity on Effective Message Design. *Communication Studies*. Vol. 65. No. 1. P. 72—95. https://doi.org/10.1080/10510974.2013.775955.

4. Betsch C., Brewer N. T., Brocard P., Davies P., Gaissmaier W., Haase N., Leask J., Renkewitz F., Renner B., Reyna V. F., Rossmann C., Sachse K., Schachinger A., Siegrist M., Stryk M. (2012) Opportunities and Challenges of Web 2.0 for Vaccination Decisions. *Vaccine*. Vol. 30. No. 25. P. 3727—3733. https://doi.org/10.1016/j.vaccine.2012.02.025.

5. Blei D. M., Ng A. Y., Jordan M. I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*. Vol. 3. P. 993—1022.

6. Broniatowski D. A., Jamison A. M., Qi S., AlKulaib L., Chen T., Benton A., Quinn S. C., Dredze M. (2018) Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health*. Vol. 108. No. 10. P. 1378—1384. https://doi.org/10.2105/AJPH.2018.304567.

7. Buller D. B., Burgoon M., Hall J. R., Levine N., Taylor A. M., Beach B., Buller M. K., Melcher C. (2000) Long-Term Effects of Language Intensity in Preventive Messages on Planned Family Solar Protection. *Health Communication*. Vol. 12. No. 3. P. 261—275. https://doi.org/10.1207/S15327027HC1203_03.

8. Burgoon M. (1995) Language Expectancy Theory: Elaboration, Explication, and Extension. In: Berger C. R., Burgoon M. (eds.) *Communication and Social Influence Processes*. East Lansing, MI: Michigan State University Press. P. 29—52.

9. Burgoon M., Pauls V., Roberts D. (2002) Language Expectancy Theory. In: Dillard J. P., Pfau M. (eds.) *The Persuasion Handbook: Developments in Theory and Practice*. Thousand Oaks, CA: SAGE Publications. P. 117—137. https://doi.org/10.4135/9781412976046.n7.

10. Cohen D. A., Scribner R. A., Farley T. A. (2000) A Structural Model of Health Behavior: A Pragmatic Approach to Explain and Influence Health Behaviors at the Population Level. *Preventive Medicine*. Vol. 30. No. 2. P. 146—154. https://doi.org/10.1006/pmed.1999.0609.

11. Đorđević J. M., Mari S., Vdović M., Milošević A. (2021) Links Between Conspiracy Beliefs, Vaccine Knowledge, and Trust: Anti-Vaccine Behavior of Serbian Adults. *Social Science & Medicine*. Vol. 277. Art. 113930. https://doi.org/10.1016/j.socscimed.2021.113930.

12. Dudina V. I., Saifulina V. O. (2023) "I Read It, I Wanted To Get Vaccinated Even Less": An Online Discourse of Vaccine Hesitancy. *Monitoring of Public Opinion: Economic and Social Changes*. No. 1. P. 279—298. https://doi.org/10.14515/monitoring.2023.1.2344. (In Russ.)
Дудина В. И., Сайфулина В. О. «Почитала, еще меньше вакцинироваться захотелось»: онлайн-дискурс вакцинной нерешительности // Мониторинг общественного мнения: экономические и социальные перемены. 2023. № 1. С. 279—298. https://doi.org/10.14515/monitoring.2023.1.2344.

13. Faasse K., Chatman C. J., Martin L. R. (2016) A Comparison of Language Use in Pro- and Anti-Vaccination Comments in Response to a High Profile Facebook*[6] Post. *Vaccine*. Vol. 34. No. 47. P. 5808—5814. https://doi.org/10.1016/j.vaccine.2016.09.029.

14. Jensen M. L., Averbeck J. M., Zhang Z., Wright K. B. (2013) Credibility of Anonymous Online Product Reviews: A Language Expectancy Perspective. *Journal of Management Information Systems*. Vol. 30. No. 1. P. 293—324. https://doi.org/10.2753/MIS0742-1222300109.

15. Jolley D., Douglas K. M. (2014) The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions. *PLOS ONE*. Vol. 9. No. 2. Art. e89177. https://doi.org/10.1371/journal.pone.0089177.

16. Kata A. (2010) A Postmodern Pandora's Box: Anti-Vaccination Misinformation on the Internet. *Vaccine*. Vol. 28. No. 7. P. 1709—1716. https://doi.org/10.1016/j.vaccine.2009.12.022.

17. Kata A. (2012) Anti-Vaccine Activists, Web 2.0, and the Postmodern Paradigm — An Overview of Tactics and Tropes Used Online by the Anti-Vaccination Movement. *Vaccine*. Vol. 30. No. 25. P. 3778—3789. https://doi.org/10.1016/j.vaccine.2011.11.112.

18. Koltsova O. Y., Alexeeva S. V., Kolcov S. N. (2016) An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. In: *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog"*. Vol. 15. No. 22. P. 277—287.

19. Larson H. J., Jarrett C., Eckersberger E., Smith D. M. D., Paterson P. (2014) Understanding Vaccine Hesitancy Around Vaccines and Vaccination from a Global Perspective: A Systematic Review of Published Literature, 2007—2012. *Vaccine*. Vol. 32. No. 19. P. 2150—2159. https://doi.org/10.1016/j.vaccine.2014.01.081.

[6] * Readers should note that Meta Platforms, Inc. and its product Facebook* are legally designated as extremist organizations and banned in the Russian Federation.

DOI: 10.14515/monitoring.2026.4.3147

20. Lee C. H., Yu H. (2020) The Impact of Language on Retweeting During Acute Natural Disasters: Uncertainty Reduction and Language Expectancy Perspectives. *Industrial Management & Data Systems*. Vol. 120. No. 8. P. 1501—1519. https://doi.org/10.1108/IMDS-12-2019-0711.

21. Liu J., Jiang H., Wang S. (2023) Physicians' Online Writing Language Style and Patient Satisfaction: The Mediator of Depth of Physician–Patient Interactions. *Healthcare*. Vol. 11. No. 11. Art. 1569. https://doi.org/10.3390/healthcare11111569.

22. MacDonald N. E. (2015) Vaccine Hesitancy: Definition, Scope and Determinants. *Vaccine*. Vol. 33. No. 34. P. 4161—4164. https://doi.org/10.1016/j.vaccine.2015.04.036.

23. Malthouse E. (2023) Confirmation Bias and Vaccine-Related Beliefs in the Time of COVID-19. *Journal of Public Health*. Vol. 45. No. 2. P. 523—528. https://doi.org/10.1093/pubmed/fdac128.

24. Mnookin S. (2011) The Panic Virus: A True Story of Medicine, Science, and Fear. New York, NY: Simon & Schuster.

25. Mohammad S. M., Turney P. D. (2013) Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*. Vol. 29. No. 3. P. 436—465. https://doi.org/10.1111/j.1467-8640.2012.00460.x.

26. Murugesan S. (2007) Understanding Web 2.0. *IT Professional*. Vol. 9. No. 4. P. 34—41. https://doi.org/10.1109/MITP.2007.78.

27. Petrov I. (2022) Anti-Vaccination Movement on VK: Information Exchange and Public Concern. In: Alexandrov D. A., Boukhanovsky A. V., Chugunov A. V., Kabanov Y., Koltsova O., Musabirov I. (eds.) *Digital Transformation and Global Society*. Cham: Springer. Vol. 1503. P. 108—121. https://doi.org/10.1007/978-3-030-93715-7_8.

28. Platonov K., Svetlov K. (2020) Conspiracy Theories Dissemination on SNS Vkontakte: COVID-19 Case. In: Chugunov A., Khodachek I., Misnikov Y., Trutnev D. (eds.) *Electronic Governance and Open Society: Challenges in Eurasia*. Cham: Springer. Vol. 1349. P. 322—335. https://doi.org/10.1007/978-3-030-67238-6_23.

29. Powell J., Pring T. (2024) The Impact of Social Media Influencers on Health Outcomes: Systematic Review. *Social Science & Medicine*. Vol. 340. Art. 116472. https://doi.org/10.1016/j.socscimed.2023.116472.

30. Recio-Román A., Recio-Menéndez M., Román-González M. V. (2023) Influence of Media Information Sources on Vaccine Uptake: The Full and Inconsistent Mediating Role of Vaccine Hesitancy. *Computation*. Vol. 11. No. 10. Art. 208. https://doi.org/10.3390/computation11100208.

31. Renaud L., Bouchard C., Caron-Bouchard M., Dubé L., Maisonneuve D., Mongeau L. (2006) A Model of Mechanisms Underlying the Influence of Media on Health Behaviour Norms. *Canadian Journal of Public Health*. Vol. 97. No. 2. P. 149—152. https://doi.org/10.1007/BF03405337.

32. Reyna V. F. (2012) Risk Perception and Communication in Vaccination Decisions: A Fuzzy-Trace Theory Approach. *Vaccine*. Vol. 30. No. 25. P. 3790—3797. https://doi.org/10.1016/j.vaccine.2011.11.070.

33. Taddy M. (2012) On Estimation and Selection for Topic Models. In: Lawrence N. D., Girolami M. (eds.) *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (AISTATS). PMLR. Vol. 22. P. 1184—1193.

34. Tokojima Machado D. F., de Siqueira A. F., Gitahy L. (2020) Natural Stings: Selling Distrust About Vaccines on Brazilian YouTube. *Frontiers in Communication*. Vol. 5. Art. 577941. https://doi.org/10.3389/fcomm.2020.577941.

35. Tserkovnaya D., Larson H. (2019) Vaccine Confidence in Russia: Digital Analysis of the Online Behavior and Attitudes of the Russian Speaking Internet Users. In: *Proceedings of the 9th International Conference on Digital Public Health*. New York, NY: Association for Computing Machinery. P. 135—136. https://doi.org/10.1145/3357729.3357763.

36. Vranceanu A.-M., Elbon M., Adams M., Ring D. (2012) The Emotive Impact of Medical Language. *HAND*. Vol. 7. No. 3. P. 293—296. https://doi.org/10.1007/s11552-012-9419-z.

37. Weston S. J., Shryock I., Light R., Fisher P. A. (2023) Selecting the Number and Labels of Topics in Topic. Modeling: A Tutorial. *Advances in Methods and Practices in Psychological Science*. Vol. 6. No. 2. P. 1—13. https://doi.org/10.1177/25152459231160105.