Discourse on Vaccination on Russian Social Media: Topics and Controversy

Konstantin Platonov
Higher school of Economics
Moscow, Russia
konplatonov@gmail.com

Kirill Svetlov
Peter the Great Saint-Petersburg Polytechnic University
Saint Petersburg, Russia
kir.svetloff@gmail.com

Viktoriia Saifulina
Saint Petersburg State University
Saint Petersburg, Russia
vosayfulina@gmail.com

***Abstract*—Public discussion on vaccination in Russia is vigorous and controversial, especially in the case of COVID-19. In conditions of extensively spreading myths, false information and rhetoric contradicting argumentation of scientific community social media became a place where opinions on vaccination collide. Participatory culture of commenting still remains to be a peculiar form of public health activism accessible to almost everyone. In this study, the data retrieved from the most popular Russian social networking platform Vkontakte was used. The raw dataset included 467888 news posts (published during 2021) from salient online communities and 538202 comments. Topic mining and modeling methods including PLSA and LDA were used to classify vaccination-related news posts in 6 groups, which differed in terms of language style, main focuses and discussed issues. The most engaging topic was "Vaccination on the ground" mainly because in contained an abundance of "obtrusive" issues. It was shown that the degree of user engagement did not significantly depend on salience of topic. In sum, it was revealed that 6.2% of comments was against vaccination, while, one and a half times less, 4.3% was in favor. Positive comments outweighed the negative ones only for topic "Russian vaccines in the World".**

I. INTRODUCTION

After the beginning of pandemic, the topic of vaccination, which had already been controversial [1], became more salient on the media agenda [2, 3]. In "infodemic" context an abundance of inconsistent information sources merged, and it has had particularly strong effect on social media [4]. The problem was exacerbated by the ability of fake news to spread faster than "true stories", making it difficult for majority of users to distinguish verified facts from fabricated ones [5]. Many countries undertake systematic measures to reduce the flow of misinformation but there are certain contradictions in this process. For instance, targeting undesired opinions using methods of state regulation and law enforcement can resemble a form of censorship, and at the same time the critique of vaccination is not always anti-vax propaganda, but can come from scientifically based concerns [2]. Still, anti-vax activism can lead to growth of echo chambers, information bubbles, false consensus and development of conspiracy theories, it also can be used for political purposes in the form of populist rhetoric [6].

The complexity and depth of arguments, complicated medical and epidemiological context, and the abundance of news coverage make the topic of vaccination provocative for commentators on social media. Today, the influence of disinformation campaigns is so strong that can be used to predict vaccine hesitancy and even mean vaccination coverage [7].

Social media provide fertile ground for spreading of various myths about vaccination. For example, in English language Twitter, the most prominent examples of misinformation include statements that vaccines can "affect DNA", "cause infertility", "be toxic", or "be used to spread tracking devices" [3]. According to the study of the Center for Countering Digital Hate, 65% of anti-vax content shared or posted on Facebook and Twitter was published by a specific group of people they called the Disinformation Dozen [8].

In present study, we used data retrieved from popular Russian social networking site VKontakte. Datasets included posts published in communities representing influential Russian media (top-20), as well as comments on these posts. The most part of research connected to this topic take into account negative and positive sentiments towards vaccination [9], and presented one is no exception. On the basis of the considered problems the following research questions and hypotheses can be formulated.

RQ1. What topics can be identified in the news related to vaccination in VKontakte communities representing the top-20 Russian media?

H1. In online communities representing the top-20 Russian media, the more a particular "vaccination-related" topic engages the audience, the more often it is presented in the overall news stream.

RQ2. What are the shares of "pro-vax" and "anti-vax" comments in the identified topics?

H2. Proportion of comments with different tones ("pro-vaccine" and "anti-vaccine") depends on topic of posts.

II. RELATED WORK

Data mining and topic modeling are among the key methods in the research related to COVID-19-related discussions on social media [10], [11]. Studies showed the diversity and multidimensional character of discourse on vaccination, COVID hesitancy and misinformation [12], [13]. Some authors demonstrated correlation between the sentiment score of tweets and the number of newly confirmed cases [14]. The others attempted to predict COVID outbreaks via social media mining and revealed strong correlation between the number of official cases and the number of relevant posts and searches [15].

Despite the risks related to the spread of misinformation some papers insisted on the importance of anti-vaxxers in

COVID discourse because of its role in “fostering civil dialogue” [16]. Anti-vaxxers use diverse arguments to push their discourse, for example, a cross-platform study of Facebook, Twitter, Instagram, and TikTok described 14 different types of argumentations, which were revealed via manual classification of subsets of comments and described by keyword sets. Topics were labeled based on main attitudes, for example, “I do not want to be vaccinated, because I have freedom of choice”, “the vaccines do not work”, and “no one is responsible for the potential side effects of the vaccine” [17]. Such approaches typically use either topic modeling or manual content analysis to separate one topic from another. Other studies revealed particular words specific to negative (towards vaccination) texts, for instance, “government, state, science, employee, risks, and also some expletives”. [11]

The research on discussion on COVID vaccination in social media, for instance, Reddit [11] and Twitter is plentiful [18]. Vkontakte is obviously a more country-specific SNS, but there are numerous articles published using it as a source, as well. The overall majority of them are devoted to questions related to spreading misinformation and COVID-related conspiracy [19], [20].

III. Data and method

In order to collect a sample of prominent news media, top-20 of media outlets were chosen according to the rating of Brand Analytics (December 2021) [21]. This is an often-used social media monitoring system that implements automated algorithms and publishes opened rankings on a regular base.

Then, corresponding VKontakte communities were selected, all of them, were fairly active with a number of subscribers from 11 thousand to 2.8 million (Table I).

TABLE I. List of Vkontakte communities used in the study

Name	Short link
RIA Novosti (RIA News)	https://vk.com/ria
RT in Russian	https://vk.com/rt_russian
Komsomol'skaya Pravda (Komsomol Truth)	https://vk.com/kpru
RBC	https://vk.com/rbc
TASS	https://vk.com/tassagency
Telekanal Tsargrad	https://vk.com/tsargradtv
Sports.ru	https://vk.com/sportsru
Lenta.ru	https://vk.com/lentaru
Izvestia (IZ.RU)	https://vk.com/izvestia
Kommersant (Merchant)	https://vk.com/kommersant
Rossijskaya Gazeta (Russian Newspaper)	https://vk.com/rgru
Gorod Moskva (Moscow City)	https://vk.com/mos
Argumenty i Fakty (Arguments and Facts)	https://vk.com/aif_ru
Federal'noe agentstvo novostej (Federal news agency)	https://vk.com/riafan
Novostnoy front (News Front)	https://vk.com/newsfront_tv
NGS.News	https://vk.com/news_ngs
Moscow's Komsomolets	https://vk.com/mk_ru
Znak (Sign)	https://vk.com/znak_com
Krasnaya Vesna (Red Spring)	https://vk.com/rossaprimaveraru
Echo of Moscow	https://vk.com/blocked_channel123

The choice of VKontakte social network was associated with its popularity among users across ages and a good representation of news communities in it.

The daily audience of Vkontakte in the first quarter of 2022 was estimated at more than 47 million users, and monthly audience was more than 100 million. In March 2022 users left more than 468 million posts and comments. Also in March, the number of communities created increased by 68% compared to the previous year [22].

Using VK API and custom Python scripts, 467888 news posts with texts were retrieved from the communities, if the text was also presented in the repost, the strings of texts were concatenated. The time frame covered the entire year 2021. Our sources proved to be fairly stable in terms of publication activity over time. Most of them published at least 1000 posts per month. Both public and private media, as well as thematic platforms with a particular focus on business, politics, or sports, were represented, but all had a considerable share of posts about vaccination.

On the next stage, 17092 relevant vaccination-related news posts were selected from database using topic modeling based on PLSA and BigARTM [23], [24]. To build the topic model, all posts underwent additional processing: 1) the most frequent words (in more than 50% of the documents) were removed from the texts, 2) the words present in less than 20 documents were removed from the texts, too. Thus, words that did not affect the meaning of the message, such as conjunctions, as well as words with errors and misprints, were filtered out.

Using BigARTM library, which implements the modern approach of Additive Regularization, a topic model was constructed for the considered corpus of texts. All texts were divided into 50 topics. The key topic with standard deviation $\sigma=0.1124$ and average share $\mu=0.0242$ was extracted. To select the texts for the topic of interest, a heuristic approach was used, which consisted in selecting messages whose share in topic exceeded $\mu+2\sigma$.

The keyword set for general target topic included the following terms: «vaccine», «vaccination», «coronavirus», «Sputnik», «COVID», «inoculating», «drug», «Russian Ministry of Health», «coronavirus», «health care», «Russia», «declare», «certificate», «report», «country», «vaccinate», «detailed», «inoculation».

The second dataset was represented by comments on these posts, 538202 comments in total, they were also obtained via VK API. Some of them turned out to be deleted, but the total number was negligible (1.1%).

There is a significant skew in the number of posts in relevant communities. (Fig. 1)

Word frequency counts were implemented for nouns, adjectives, and verbs. The most frequent terms in the texts were, as might be expected, such typical words as "vaccine", "coronavirus", "vaccination", "Sputnik" (Sputnik V is the most important Russian vaccine from COVID-19) and "COVID"; terms related to politics, science, health care and its institutions were also listed in the top-30. The most frequent words list also included proper names associated with the Russian vaccination agenda, such as "Gamaleya" (Russian vaccine Sputnik V was developed in The National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya) and "Ginzburg" (Alexander Ginzbirg is the head of this Center). (Fig. 2)

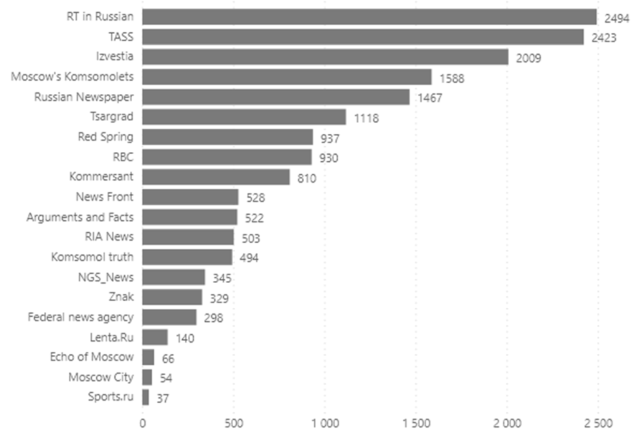

Fig. 1. Distribution of the number of posts in the filtered array (iteration 1)

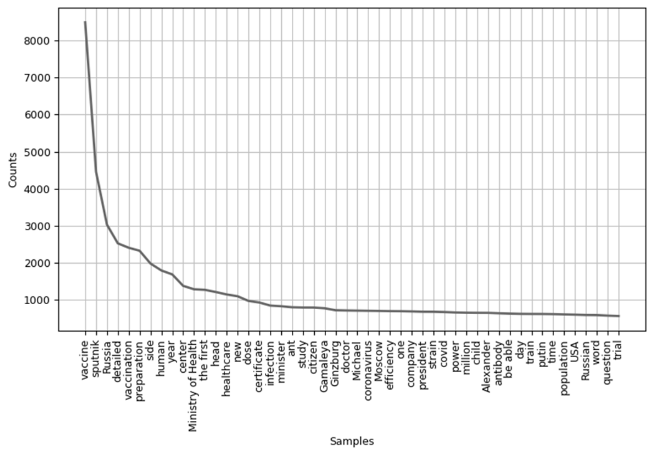

Fig. 2. Top 50 words in vaccination news (iteration 1)

The first iteration showed that the filtered data included a small part of news that the algorithm classified as vaccination-related, while they were not. Therefore, the second iteration was necessary. To filter out individual news stories related to urban incidents, politics, economics and other topics, 50 words which are typical in the topic of coronavirus and vaccination were selected from the top 500 frequent words. Further, only posts with at least one of these terms were retained which significantly improved the quality of the dataset. The number of posts was reduced from 17092 to 14816.

The wordlist included the following words: «vaccine», «coronavirus», «vaccination», «Sputnik», «covid», «coronavirus», «dose», «certificate», «vaccinate», «Ginzburg», «to vaccinate», «drug», «research», «Gamaleya», «strain», «antibody», «Astrazeneca», «Pfizer», «medical», «clinical», «recover», «doctor», «health care», «Epivaccorona», «immunity», «pandemic», «component», «revaccination», «Omicron», «infection», «virus», «medic, «get sick», «PCR», «Moderna», «contraindication», «get infected», «epidemiology», «morbidity», «vaccinate», «anti-vaccine», «unvaccinated», «wave», «Procenko».

Outlets' interest in vaccination topic varied over time: for example, during most months TASS produced the most posts, while in May, June and July the most active source was RT. In addition, there were certain temporal trends such as spikes in January, March, June/July, and November.

Python libraries were used to analyze the presence of different topics in texts. Basic dataset consisted of 14816 posts. The average length of texts in this corpus was 14 words and IQR was 7-17 words. Duplicates and near-duplicates were saved in order to take into account the original distribution over sources (the number of unique texts was 14721). Below is the analysis algorithm.

1) Creation of dataframe ("pandas" library data type) [25].
2) Tokenization of texts and posts with nltk library [26].
3) Parsing and filtering with pymorphy2 [27] library. At this stage, insignificant words were removed from the texts, preprocessing was carried out, only nouns and adjectives (short and full), as well as proper names and abbreviations were kept for the following steps. The cleaning of stop words was carried out using the stopwords set from nltk library, which was expanded with additional terms.
4) Creation of bigrams, topic modeling with genism library [23].
5) Visualization of the LDA model with LDAvis library [28].

Because further clustering texts belonging to a certain topic is complicated by itself, the number of topics for (the second) topic identification was defined empirically. To select the optimal number of topics, a coherence calculation was implemented, and the best test scores were obtained for 5, 6, 7, 9 topics with values in the range from 0.45 to 0.48 (Fig. 3). The best interpretability and distance between topics (intertopic distance) was achieved on 6 topics.

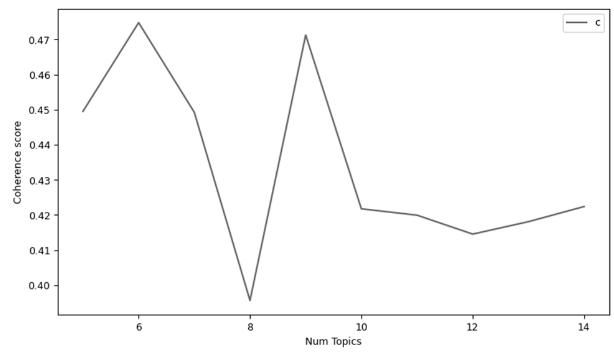

Fig. 3. Coherence estimation for LDA models across the number of topics varying from 5 to 15.

IV. Results

A. Topics in posts

By choosing the parameters of the model, it was possible to further improve the overall coherence coefficient. As a result, the coherence coefficient value of 0.51 for 6 topics was obtained. The differences between average lengths of texts for different topics were not significant. Each topic was described through the top-10 frequent words and the corresponding weight coefficients. The higher the coefficient, the higher the share of a word in the topic. The topics were given fairly conventional names to distinguish them from each other.

Topic 1. "Federal vaccination agenda". The topic describes primarily the Russian vaccination management agenda as a planned process in the context of politics and public health. The key actors in this topic were the Ministry of Health and Rospotrebnadzor (Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing).

0.039*Russia + 0.038*coronavirus + 0.027*vaccination + 0.019*Ministry of Health + 0.016*year + 0.015*health care + 0.015*RF + 0.013*Minister + 0.013*head + 0.013*Murashko

Topic 2. "Vaccination in practice". The topic described the situation of vaccine research, use, and development in Russia and (to a lesser extent) the world in the context of science, epidemiology, and medical practice. The main "representatives" were Russian vaccines, such as Kovivac, Epivakkorona, Sputnik V, as well as the Gamaleya Research Institute and its head Alexander Gintsburg, Vector Institute (State Research Center of Virology and Biotechnology), in which Epivakkorona vaccine was created.

0.067*vaccine + 0.034*center + 0.030*Sputnik + 0.029*coronavirus + 0.021*study + 0.018*strain + 0.017*Gamaleya + 0.017*Gintsburg + 0.016*test + 0.016*preparation

Topic 3. "Vaccination on the ground". This topic described various stories about the properties and effectiveness of vaccines, contraindications and side effects, the vaccination procedure and "vaccination in the field", including regional and local issues (often in Moscow). This topic focused more on the actual interaction between the medic, the patient and the vaccine than the previous.

0.079*vaccination + 0.033*inoculation + 0.027*coronavirus + 0.018*detailed + 0.017*person + 0.014*doctor + 0.013*child + 0.011*obligatory + 0.009*region + 0.009*Moscow

Topic 4. "Formal side of vaccination". This topic involved everything about vaccination documents, restrictions and tolerances, state policies on border crossings during a pandemic, vaccination certificates, QR-codes, COVID-19 tests, and supervision.

0.040*certificate + 0.037*vaccination + 0.022*coronavirus + 0.018*test + 0.012*vaccination + 0.012*which + 0.011*code + 0.009*power + 0.009*citizen + 0.009*availability

Topic 5. "Foreign vaccination agenda". In general, this topic focused on vaccination in USA, WHO activity, Pfizer, AstraZeneca, and Johnson & Johnson vaccines. It mostly contained discussing their effectiveness and side effects.

0.030*vaccine + 0.024*WHO + 0.021*coronavirus + 0.020*vaccination + 0.018*detailed + 0.016*organization + 0.013*health care + 0.012*USA + 0.012*drug + 0.011*side effect

Topic 6. "Russian Vaccines in the World". The topic focused on the registration and approbation of Russian vaccines (mainly Sputnik V) in other countries.

0.090*vaccine + 0.056*satellite + 0.040*Russian + 0.029*detailed + 0.028*coronavirus + 0.020*country + 0.018*drug + 0.012*dose + 0.011*Russia + 0.010*company

B. User engagement

Engagement rate for each news outlet post was calculated using presented formula.

$$Engagement\ rate = \frac{(likes + comments)}{views} * 1000$$

Topics significantly differed in audience involvement. Topic "Vaccination on the ground" (Key words: «vaccine», «Sputnik», «citizen», «vaccination», «coronavirus», «clinic», «detailed») dominated in user engagement, probably, because it was understandable and familiar to wide audience. Apparently, for similar reasons, the average commenting and "liking" rates were lower for topics "Foreign vaccination agenda" and "Russian Vaccines in the World. (Fig. 4)

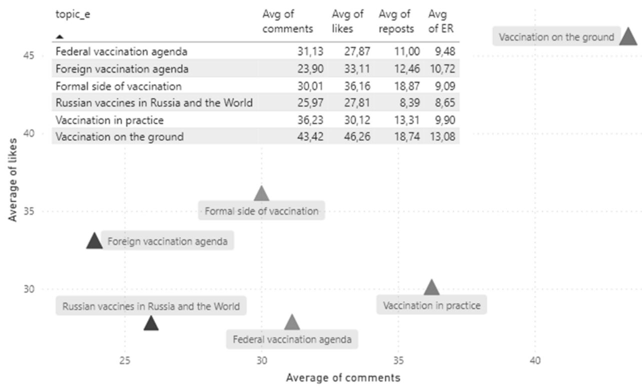

Fig. 4. Descriptive statistics for topics

The average number of reposts was higher for topics "Vaccination on the ground" and "Formal side of vaccination", probably because they most often contained information which could be useful for ordinary users, which made people want to share it. The proportions of topics for different media varied. For example, the Moscow mayor official community did not write about "Russian vaccines in the World", the Federal News Agency covered "Foreign vaccination agenda" less frequently than other communities, while the largest and most active media outlets had a more uniform distribution of topic presence.

First hypothesis about the ability of the most common topics to provoke the greatest response was tested by analyzing how engagement rates varied by topics and their prevalence. According to engagement rates there was no evidence to prove the hypothesis. The most engaging topic was "Vaccination on the ground". This topic ranked only third out of seven ones. In turn, the least engaging topic was the largest in terms of number of posts. (Fig. 5)

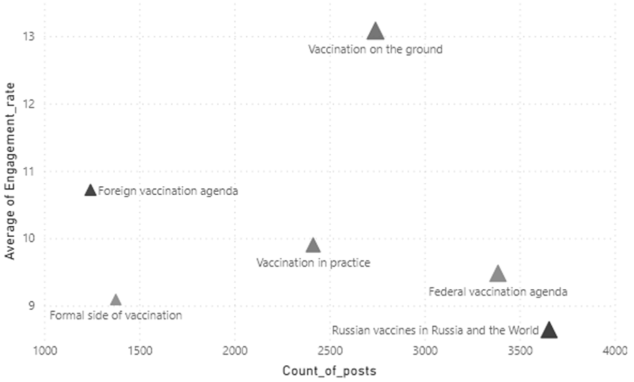

Fig. 5. Distribution of the average engagement rate and the number of posts by topic.

Whilst there was no direct and stable relationship between topic prevalence and average engagement rate, but different topics did vary in their engagement rates. The pair-wise Tukey HSD test (implemented with “statsmodels” library for Python) also did not reveal any pattern in pairwise comparisons. The total number of posts in pair-wise comparisons did not determine any significant differences in engagement rate. (Table II)

TABLE II. DIFFERENCES IN ENGAGEMENT RATES BETWEEN GROUPS

Group1	Group2	Module of the differences in ERs	Number of posts for group1	Check for hypothesis of group differences
Russian vaccines in the World	Federal vaccination agenda	0.9226	3655	True
Russian vaccines in the World	Formal side of vaccination	0.5967	3655	False
Federal vaccination agenda	Formal side of vaccination	0.3259	3386	False
Vaccination on the ground	Foreign vaccination agenda	2.2684	2742	True
Vaccination on the ground	Vaccination in practice	3.2988	2742	True
Vaccination on the ground	Russian vaccines in the World	4.6526	2742	True
Vaccination on the ground	Federal vaccination agenda	3.7300	2742	True
Vaccination on the ground	Formal side of vaccination	4.0559	2742	True
Vaccination in practice	Russian vaccines in the World	1.3538	2414	True
Vaccination in practice	Federal vaccination agenda	0.4312	2414	False
Vaccination in practice	Formal side of vaccination	0.7571	2414	False
Foreign vaccination agenda	Vaccination in practice	1.0304	1243	False
Foreign vaccination agenda	Russian vaccines in the World	2.3842	1243	True
Foreign vaccination agenda	Federal vaccination agenda	1.4616	1243	True
Foreign vaccination agenda	Formal side of vaccination	1.7875	1243	True

For a general overview, the main features of each topic in the filtered dataset, can be presented as weighted (by views) metrics for the median number of comments, reposts and likes: ERc, ERr, ERl, calculated by analogy with Engagement rate. (Table III)

TABLE III. METRICS FOR TOPICS (MEDIANS)

Topic	ERc	ERl	ERr	Views	ER
Vaccination on the ground	2.94	3.28	1.23	4507	7.37
Foreign vaccination agenda	1.71	3.23	0.92	4469	5.95
Vaccination in practice	2.69	3.10	0.93	4975	6.49
Russian vaccines in the World	1.96	3.32	0.49	5318	6.17
Federal vaccination agenda	2.55	3.02	0.83	4617	6.19
Formal side of vaccination	1.94	2.90	1.24	4230	5.71

To test our second hypothesis about the distribution of comments in support and against vaccination among topics, 10000 comments were randomly selected from the filtered dataset and coded manually as pro- and anti-vaccination. Only comments expressing explicit opinion received codes. Cohen's Kappa value of 0.64 was achieved which could be classified as good or substantial [29]. 6.2% of the comments in subsample were against vaccination, and 4.3%, 2.5 times less, were pro-vaccination.

Chi-square test (implemented using “scipy.stats” Python library) $\chi^2 = 17.2$ and p-value = 0.04. Hence, the topic and the tone of a given post (“pro or against”) were not independent (Table 4). The topic “Russian vaccines in the World” was the only one for which the share of positive comments was higher than negative ones. The topic “Vaccination in practice”, on the contrary, had the higher difference between “against” and “in favor” (33%). However, crosstab “topic-tone” Cramer’s produced V = 0.13, showed a weak link between the measures. Thus, we considered the second hypothesis partly accepted.

TABLE IV. SHARES OF «AGAINST AND PRO» COMMENTS FOR DIFFERENT TOPICS

Topic	Count “against”	Count “pro”	Total	Share “against”	Share “pro”
Vaccination on the ground	198	111	309	0.641	0.359
Foreign vaccination agenda	37	23	60	0.617	0.383
Vaccination in practice	120	61	181	0.663	0.337
Russian vaccines in the World	79	86	165	0.479	0.521
Federal vaccination agenda	124	100	224	0.554	0.446
Formal side of vaccination	46	35	81	0.568	0.4

C. “Pro” and “against” argumentation

The argumentation against vaccination in comments was quite diverse and could be grouped into several types.

1) Statements about the lack of trust to actions and institutions of the state which is accused of manipulation and "covering up the truth".
2) Statements about "big pharma conspiracy" to get more revenue.
3) Statements on "vaccination as a form of experiments on humans".
4) Reasoning related to the interpretation of statistics. For example, some anti-vaxxers claim a link between the growth of the share of the vaccinated and the share of the infected.
5) Critique of the management of vaccination. Some commentators opposed mass vaccination during the pandemic in general.
6) Statements on lack of "reliable data" about effectiveness and safety of vaccines.
7) Critique of Russian science and healthcare and doubts about the ability to develop a proper vaccine.
8) Opinions about uselessness of vaccination such as "it is better to get sick and develop natural immunity".
9) Conspiracy on the ability of the vaccine to "influence DNA" and "provoke deceases".
10) Critique of Gamaleya Research Institute as a vaccine developer.
11) Mentioning the fact that Russian vaccine is not accepted by WHO.
12) Conspiracies claiming, for instance, that "vaccination is chipping, institutionalized segregation, instrument of state control".

The argumentation in favor of vaccination could be summarized in several groups.

1) Mentioning the idea that "to get vaccinated" means to take care of yourself and others around you in pandemic (as an act of civil liability).
2) References to the proven efficiency of vaccine in terms of reducing the risk of development of complications associated with COVID-19.
3) Critique of anti-vaxxers as "criminals who foster spreading of COVID by their ignorance".
4) References to positive personal or other person's vaccination experience. Or, alternatively, references to experiencing more severe cases of COVID.
5) Statement about any government's responsibility and interest in health of its citizens; claiming any critique of vaccination as unreasonable
6) Claims of equality of effectiveness among vaccines; explanation of the controversy as a "war for market outlet".
7) Statement about the vaccines' being the only accessible way of battling COVID-19, since no effective treatment has been proposed.

V. Discussion

The findings showed that the discussion around vaccination in Russian social media has a complex structure including wide variety of arguments. The increased engagement rate in topics "Vaccination in practice" and "Vaccination on the ground" is related to close relation to the actual practical experience of audiences in relation to vaccination. These two topics also produced the largest differences between the share of positive and negative comments. The topic "Russian vaccines in the World" was the least interesting for users (the lowest average engagement rate), however, this topic contained the highest number of posts, which might indicate a certain form of agenda-setting of this topic, related to the promotion of Russian vaccines.

Limitations

Our choice of communities relied on one rating by Brand analytics, which can be rather "subjective". The sample for sentiment coding was limited by 10000, it could not be automated because of the complexity debated questions. It also should be mentioned that automatic classification using LDA and PLSA has its own constraints [30, 31].

VI. Conclusion

Presented algorithm involving two-step text clustering proved to be an effective tool to define and describe particular agenda represented by social media landscape in Russian prominent news media online-communities.

Using topic mining, the structure of subtopics in the main topic related to vaccination was described (6 topics in total: "Federal vaccination agenda", "Vaccination in practice", "Vaccination on the ground", "Formal side of vaccination", "Foreign vaccination agenda", "Russian vaccines in the World"). The main rationale for this distinction was in the combination of level of the agenda (international, federal, local) and focus varying from vaccination in terms of epidemiology to inoculation as a technical procedure. Similar resulting classifications can be found in another research in this domain [32].

In our dataset, the presence of topic (the number of news posts) did not predict the level of user engagement. The distribution of positive and negative comments did not show any repetitive pattern across subtopics of news posts. Therefore, according to engagement rates there was no evidence to prove the H1.

While the landscape of public opinion, as pictured in users' responses in social media communities, appears fragmented and incoherent, we identified general trends. Particularly, negative opinions about the vaccination were voiced more frequently in all but one topic ("Russian vaccines in the World"). Posts containing practical information relating to vaccination (news post topic "Vaccination on the ground") received the highest engagement. Thus, we considered the H2 partly accepted.

Acknowledgment

The research was supported by RSF (project No 22-18-00261).

References

[1] S. L. Benoit and R. F. Mauldin, "The "anti-vax" movement: a quantitative report on vaccine beliefs and knowledge across social media", *BMC public health*, 21(1), 2021, pp. 1-11.

[2] R. Armitage, "Online 'anti-vax'campaigns and COVID-19: censorship is not the solution". *Public Health*, 190, e29, 2021, doi: 10.1016/j.puhe.2020.12.005

[3] K. Hayawi, S. Shahriar, M. A. Serhani, I. Taleb, and S. S. Mathew, “ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection”, *Public health*, 203, 2022, pp. 23-30.

[4] Edelman Trust Barometer 2020, https://www.edelman.com/trust/2020-trust-barometer

[5] P. Dizikes, “On Twitter, false news travels faster than true stories”, MIT, http://news.mit.edu/2018/study-twitterfalse-news-travels-faster-true-stories-0308

[6] N. Rietdijk, “Radicalizing Populism and the Making of an Echo Chamber: The Case of the Italian Anti-Vaccination Movement”, *Krisis*, 2021, 41(1), pp. 114-134.

[7] N. Puri, E. A. Coomes, H. Haghbayan, and K. Gunaratne, “Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases”, *Human vaccines & immunotherapeutics*, 16(11), 2020, pp. 2586-2593.

[8] Center for Countering Digital Hate, “The disinformation dozen: Why platforms must act on twelve leading online anti-vaxxers”, https://www.counterhate.com/_files/ugd/f4d9b9_b7cedc0553604720 b7137f8663366ee5.pdf

[9] E. Karafillakis, S. Martin, C. Simas, K. Olsson, J. Takacs, S. Dada, and H. J. Larson, “Methods for social media monitoring related to vaccination: systematic scoping review”, *JMIR public health and surveillance*, 7(2), 2021, e17149, doi: 10.2196/17149

[10] A. Goel and L. Gupta, “Social media in the times of COVID-19”, *Journal of clinical rheumatology*, doi: 10.1097/RHU.0000000000001508

[11] C. A. Melton, O. A. Olusanya, N. Ammar, and A. Shaban-Nejad, “Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence”, *Journal of Infection and Public Health*, 14(10), 2021, pp. 1505-1512.

[12] L. C. Jiang, T. H. Chu, & M. Sun, “Characterization of vaccine tweets during the early stage of the COVID-19 outbreak in the United States: topic modeling analysis”, *Jmir Infodemiology*, 1(1), 2021, e25636, doi: 10.2196/25636

[13] T. Nuzhath, S. Tasnim, R. K. Sanjwal, N. F. Trisha, M. Rahman, S. F. Mahmud, ... and M. M. Hossain, “COVID-19 vaccination hesitancy, misinformation and conspiracy theories on social media: A content analysis of Twitter data”, 2020, doi:10.31235/osf.io/vc9jb

[14] J. G. Shim, K. H. Ryu, S. H. Lee, E. A. Cho, Y. J. Lee, and J. H. Ahn, “Text mining approaches to analyze public sentiment changes regarding COVID-19 vaccines on social media in Korea”, *International Journal of Environmental Research and Public Health*, 18(12), 2021, 6549, doi:10.3390/ijerph18126549

[15] S. Yousefinaghani, R. Dara, S. Mubareka, and S. Sharif, “Prediction of COVID-19 waves using social media and Google search: a case study of the US and Canada”, *Frontiers in public health*, 9, 656635, doi:10.3389/fpubh.2021.656635

[16] R. Boodoosingh, L. O. Olayemi, and F. A. L. Sam, “COVID-19 vaccines: Getting Anti-vaxxers involved in the discussion”, *World development*, 136, 2022, 105177, doi: 10.1016/j.worlddev.2020.105177.

[17] D. Wawrzuta, J. Klejdysz, M. Jaworski, J. Gotlib, and M. Panczyk, “Attitudes toward COVID-19 Vaccination on Social Media: A Cross-Platform Analysis”, *Vaccines*, 10(8), 2022, 1190, doi:10.3390/vaccines10081190

[18] H. Yin, X. Song, S. Yang, and J. Li, “Sentiment analysis and topic modeling for COVID-19 vaccine discussions”, *World Wide Web,* 25(3), 2022, pp. 1067-1083.

[19] D. V. Boguslavsky, K. S. Sharov, and N. P. Sharova, “Counteracting conspiracy ideas as a measure of increasing propensity for COVID-19 vaccine uptake in Russian society”, *Journal of global health,* 2022, 2, doi: 10.7189/jogh.12.03013

[20] K. S. Sharov, “Adaptation to SARS-CoV-2 under stress: Role of distorted information”, *European journal of clinical investigation*, 50(9), 2022, e13294, doi: 10.1111/eci.13294

[21] Brand Analytics. Top-100 viral Russian media outlets, 18.01.2022, https://br-analytics.ru/blog/top-100-media-decemder-2021/

[22] Vkontakte. VKontakte summed up the results of the first quarter of 2022, https://vk.com/press/q1-2022-results

[23] Gensim python library, https://radimrehurek.com/gensim/

[24] K. Vorontsov and A. Potapenko, “Additive regularization of topic models”, *Machine Learning*, 101(1), 2015, pp. 303-323.

[25] Pandas python library, https://pandas.pydata.org/

[26] NLTK python library, https://www.nltk.org/

[27] Pymorphy2 python library, https://github.com/kmike/pymorphy2

[28] PyLDAvis python library, https://pypi.org/project/pyLDAvis/

[29] D.V. Cicchetti, S.A. Sparrow, “Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior”, *American Journal of Mental Deficiency*, 86, 1981, pp.127-137.

[30] Y. Chen, H. Zhang, R. Liu, Z. Ye, and J. Lin, “Experimental explorations on short text topic mining between LDA and NMF based Schemes”, *Knowledge-Based Systems,* 163, 2019, pp. 1-13.

[31] X. Yan, J. Guo, Y. Lan, and X. Cheng, “A biterm topic model for short texts”, *In Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 1445-1456.

[32] J. C. Lyu, E. Le Han, and G. K. Luli, “COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis”, *Journal of medical Internet research*, 23(6), 2021, e24435, doi: 10.2196/24435