



Факультет социальных наук

20.04.2023

Деструктивная коммуникация в русскоязычном политическом сегменте Telegram 2021-2022 гг.

Предзащита выпускной квалификационной работы

Научный руководитель: Ахременко А.С.

Выполнил: Соловьев В.А.



## Актуальность исследования

- Деструктивные высказывания, демонстрирующие враждебность по отношению к другим людям, могут выступать в качестве проявления аффективной поляризации (Druckman et al., 2018).
- Аффективной поляризации характеризуется увеличением враждебности по отношению к противоположной группе (Алескеров и Губенко, 2016).
- Согласно «Global Risks Report 2023», опубликованному World Economic Forum, аффективная поляризация - один из десяти основных рисков, которые могут препятствовать прогрессу общества.



## Telegram в России и мире

### Мир

- Отсутствие модераторов и деплатформинга (Rogers, 2020).
- Место сосредоточения альтернативных мнений (Rogers, 2020).
- Генератор деструктивных сообщений (Chipidza, 2022).

### Россия

- Быстро растущее социальное медиа. В 2022 году наибольший приток аудитории среди других соц медиа (ВЦИОМ, 2022).
- Высокий уровень доверия к информации. Уровень доверия вырос по сравнению с телевидением и другими социальными медиа (GroupM, 2022).
- Место для политических дискуссий (Ляховенко, 2022).



## Деструктивная коммуникация

### Понимание в русскоязычной литературе

- Деструктивная коммуникация представляет собой тип общения, направленного на сознательное преднамеренное причинение собеседнику морального и физического вреда (Волкова и Панченко, 2016).
- Деструктивная коммуникация - это сообщения, выходящие за нормы общения социальной группы (Буряковская, 2019).

### Понимание в англоязычной литературе

- Синонимом словосочетания деструктивная коммуникация выступает вредоносный контент (harmful content) (Dementeva, 2022).
- Вредоносный контент включает в себя такие виды речевого поведения, как агрессия, язык вражды и оскорбления (Smetanin, 2020).



## Основные теоретические предпосылки

- **Теория повестки дня.** Она предполагает, что средства массовой информации оказывают значительное влияние на то, какие темы и вопросы общественность считает важными, путем выборочного освещения и акцентирования определенных вопросов (MacCombs & Shaw). Исходя из этой теории мы предполагаем, что повышение деструктивной коммуникации будет происходить лишь в определенных темах, а не во всех сообщениях.
- **Спираль молчания.** Она предполагает, что люди не решаются выразить свое мнение, если чувствуют, что находятся в меньшинстве. Это связано с тем, что люди боятся социальной изоляции и отторжения, что заставляет их хранить молчание и подчиняться мнению большинства (Ноэль-Нойманн). Исходя из этой теории мы предполагаем, что деструктивная коммуникация должна колебаться на определенном уровне, а не повышаться.



## Проблема и исследовательский вопрос

**Проблема.** Деструктивная коммуникация, с одной стороны, согласно спирали молчания держаться вокруг постоянного значения, поскольку она наносит вред обществу, а с другой стороны, исследователи наблюдают рост такого общения в сети интернет (Chipidza, 2022; Rogers, 2021).

**Исследовательский вопрос.** Какие факторы способствуют изменению деструктивной коммуникации в русскоязычном политическом сегменте Telegram в 2021-2022 гг.?



## Цели и задачи

**Цель работы.** Выявление деструктивной коммуникации в русскоязычных политических Telegram-каналах за 2021-2022 гг.

**Задачи:**

- Проанализировать литературу по проблематике деструктивной коммуникации в социальных медиа.
- Измерить деструктивную коммуникацию.
- Сравнить степень деструктивной коммуникации в 2021-2022 гг.
- Провести ручную разметку данных для 2021 и 2022 годов.
- Выявить тематики сообщений в русскоязычных политических Telegram-каналах.
- Класстеризировать Telegram-каналы на основе сетевой структуры.
- Проинтерпретировать полученные результаты.



## Гипотезы

- $H_1$ : в результате, произошедших в 2022 году, общественно значимых событий, повысилась степень деструктивной коммуникации внутри русскоязычного политического сегмента Telegram.
- $H_2$ : общественно значимые события 2022 года вызвали качественную трансформацию деструктивной коммуникации в русскоязычном политическом сегменте Telegram.
- $H_3$ : повышение степени деструктивной коммуникации в 2022 году вызвано тематикой сообщений.
- $H_4$ : повышение степени деструктивной коммуникации в 2022 году вызвано политической поляризацией.





# Исследование деструктивной коммуникации в Telegram

## Эмпирическая база:

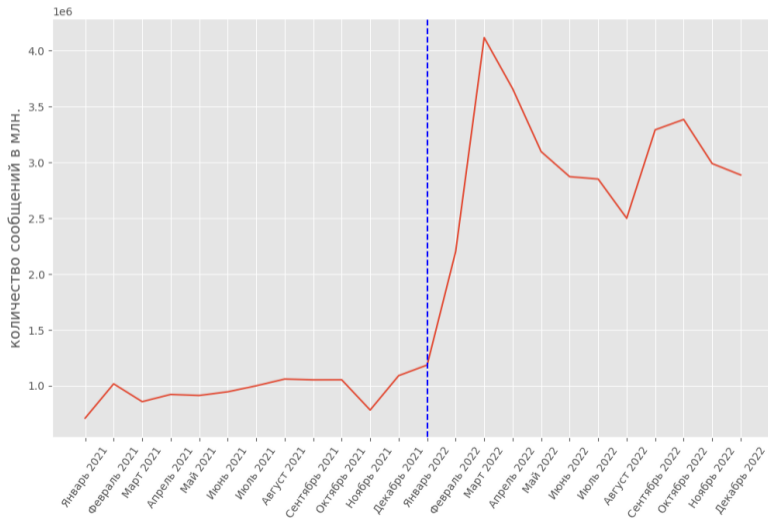
- сообщения из Telegram-каналов за 2021-2022 гг., собранные методом снежного кома, спустя 3 итерации алгоритм был остановлен (Barabashi & Albert, 1999);
- начальные точки для анализа: К. Собчак (оппозиционный журналист) и В. Соловьёв (провластный журналист). Точки взяты исходя из политической поляризации российского общества (Стукал и др., 2022);

## Инструменты для анализа данных:

- бинарный классификатор деструктивной коммуникации (Smetanin, 2020) на основе Conversational ruBERT (Kuratov & Arkhipov, 2019);
- мультиклассовый классификатор сенситивных тем (Babakov et al., 2021);
- тематическое моделирование (Grootendorst, 2022).

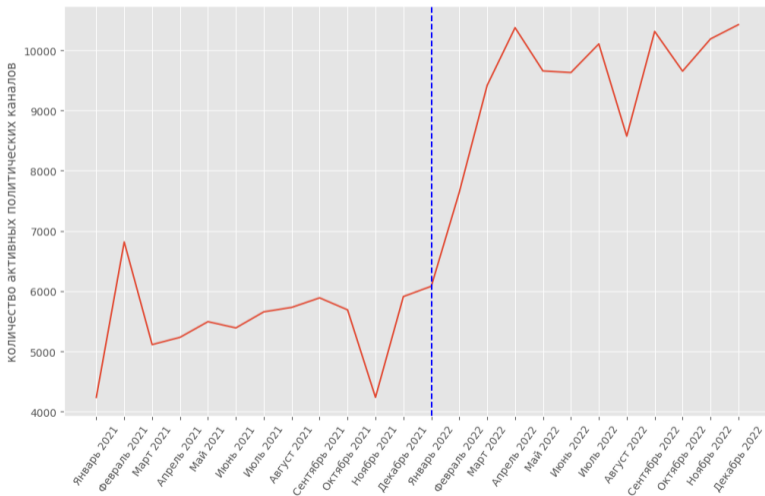


## Данные после предварительной обработки (распределение сообщений)





## Данные после предварительной обработки (распределение активных каналов)





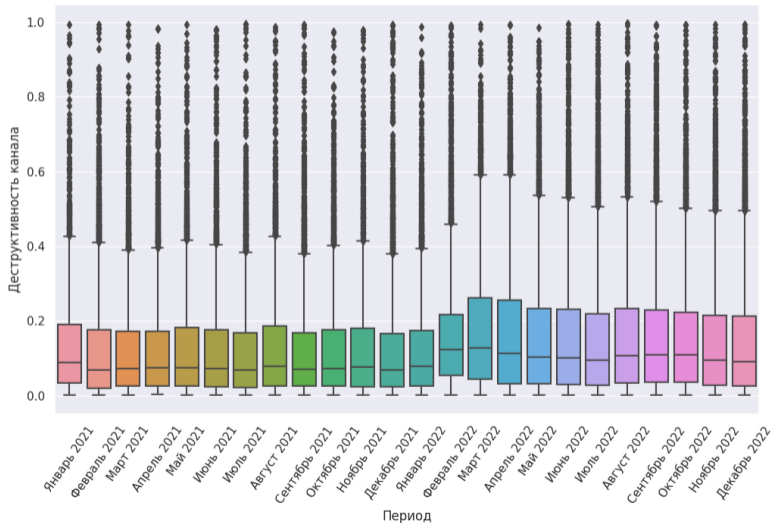
## Измерение деструктивной коммуникации ( $H_1$ )

### Модель

- На основе большой языковой модели «Conversational RuBERT», был дообучен классификатор для выявления деструктивной коммуникации в сообщениях.
- Данные для обучения были получены с 2ch.hk и ok.ru.
- Выходной слой представлен сигмоидной функцией.
- Качество модели более 92% на метриках recall, precision, F1.



# Результаты измерения деструктивной коммуникации по каналам ( $H_1$ )





## МНК регрессия ( $H_1$ )

- Спецификация:

$$\text{DesCom} = \beta_0 + \beta_1 \cdot \text{Year} + \beta_2 \cdot \text{Old}$$

- **Зависимая переменная** ( $\text{DesCom}_c$ ) - среднее значение деструктивной коммуникации канала в месяц, ( $\text{DesCom}_m$ ) - значение деструктивности сообщения.
- **Независимая переменная** ( $\text{Year}$ ) - год (2021/2022).
- **Контрольная переменная** ( $\text{Old}$ ) - показывает об активности канала в 2021 году.



## Интерпретация результатов проверки гипотезы 1 ( $H_1$ )

- В 2022 году увеличилась деструктивная коммуникация как в Telegram-каналах, так и в их сообщениях.
- Сообщения в 2022 году стали на 10% деструктивнее, а каналы на 2%.
- Сообщения из каналов, которые существовали до 2022 года, выражают в два раза меньшую деструктивность, чем остальные.



## Качественное измерение деструктивной коммуникации (H<sub>2</sub>)

- Были случайно отобраны по 500 сообщений за 2021 и 2022 годы, которые выходят за  $3Q + 1.5 * IQR$  (Inter-Quartile Range) 2021 года (слайд 15).
- Происходила ручная разметка. В роли разметчика выступал докладчик.
- Был составлен рубрикатор со следующими категориями:
  - язык вражды;
  - группа языка вражды;
  - оскорбления;
  - нецензурная лексика;
  - государство (Россия или нет);
  - исторические события (до 2000 года) и религия;
  - офлайн преступления и вооруженные конфликты.





## Интерпретация результатов тестирования гипотезы 2 (H<sub>2</sub>)

- Шансы встретить хейт-спич, нецензурную лексику и оскорбления повышаются в 2022 году.
- Таким образом, данный факт говорит о том, что происходит изменение деструктивной коммуникации не только в количественном плане, но и в качественном.



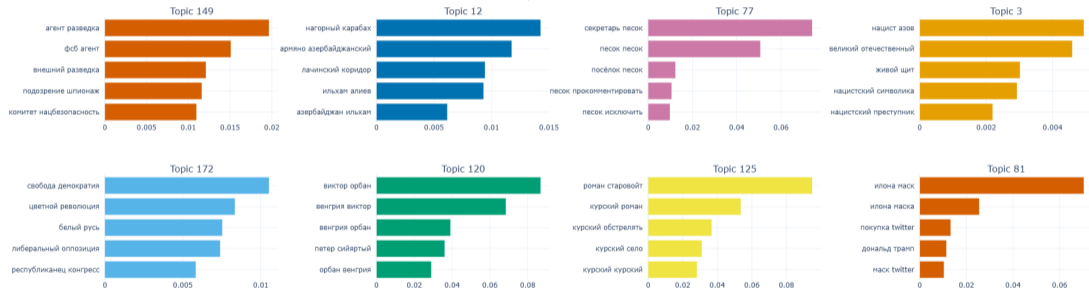
## Выявление тематик обсуждение (H<sub>3</sub>)

- BERTopic - модель для тематического моделирования с высоким уровнем согласованности тем, но ее недостаток в том, что она не выявляет нечеткие множества.
- Данные. Были взяты 200 тыс. случайных сообщений из каждого месяца, потом из этой выборки были случайно выбраны 250 тыс. для обучения BERTopic, такое решение принято исходя из имеющихся в нашем расположении вычислительных мощностей.
- После обучения модель нашла 180 тем.
- С помощью MNK мы определили, что, если сообщения относятся к одной из 79 тем (можно ознакомиться в приложении к диплому), то его деструктивность возрастает.
- 79 тем группируются в 4 больших группы: иностранные государства, публичная политика (внутренняя политика России), СВО, не интерпретируемые.



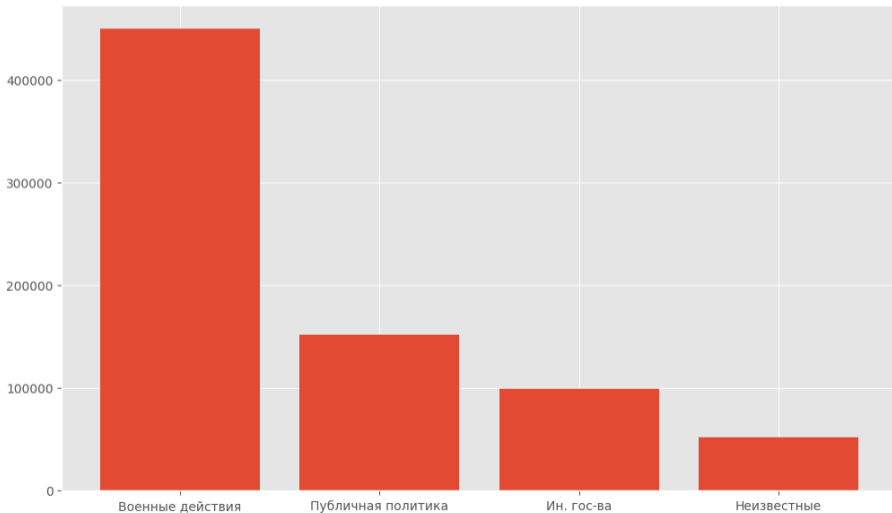
# Случайная выборка из деструктивных тем (H<sub>3</sub>)

Topic Word Scores





## Распределение сообщений по деструктивным темам (H<sub>3</sub>)





## Модель (H<sub>3</sub>)

- Спецификация:

$$\text{DesCom} = \beta_0 + \beta_1 \cdot \text{Year} + \beta_2 \cdot \text{DesTopic}_1 + \dots + \beta_6 \cdot \text{DesTopic}_6$$

- Зависимая переменная - деструктивность сообщения в выборках.
- Независимая переменная - год (Year) (2021/2022), деструктивная тема (DesTopic), где
  - DesTopic<sub>MA</sub> - военные действия;
  - DesTopic<sub>public</sub> - публичной политики,
  - DesTopic<sub>foreign</sub> - иностранные государства;
  - DesTopic<sub>unknown</sub> - неинтерпретируемые темы.



## Результаты тестирования гипотезы 3 ( $H_3$ )

Таблица 3. Деструктивность и темы сообщений

	DesCom I	DesCom II	DesCom III
Intercept	0.114*** (0.000)	0.114*** (0.000)	0.114*** (0.000)
Year	0.056*** (0.000)	0.023*** (0.000)	0.025*** (0.000)
DesTopic		0.100*** (0.000)	
DesTopic <sub>MA</sub>			0.117*** (0.000)
DesTopic <sub>public</sub>			0.071*** (0.000)
DesTopic <sub>foreign</sub>			0.063*** (0.000)
DesTopic <sub>unknown</sub>			0.083*** (0.000)
N	4.7*10 <sup>6</sup>	4.7*10 <sup>6</sup>	4.7*10 <sup>6</sup>



## Интерпретация результатов тестирования гипотезы 3 (H<sub>3</sub>)

- Деструктивные темы повышают деструктивную коммуникации, при этом уменьшается коэффициент при «Year» в два раза. Таким образом, тематика сообщений способна объяснить повышение степени деструктивной коммуникации.
- Самыми деструктивными темами являются: военные действия, неинтерпретируемая тематика, публичная политика и иностранные государства.



## Измерение поляризации через сетевую структуру (H<sub>4</sub>)

- На графе из гиперссылок происходил поиск групп с помощью жадного алгоритма поиска сообществ в сети (greedy modularity communities) (Clauset et al., 2004).
- Нахождение меры схожести между группами 2021 и 2022 года происходит по метрике Жаккара ( $\frac{|A \cap B|}{|A \cup B|}$ ).
- Для определения политической направленности выявленного сообщества был создан коэффициент соответствия, который считается следующим образом:

$$k = \frac{|\{\text{channel} : \text{channel} \in \text{community}\}|}{20}$$

где channel - 20 Telegram-каналов (10 личных и 10 коллективных оппозиционных/провластных каналов, которые отбирались экспертами), community - обнаруженные сообщества.





## Результаты выявления сообществ в 2021 и 2022 гг. (H<sub>4</sub>)

Таблица 4. размеров групп по годам.

Сообщества	2021	2022	Жаккара	k <sub>2021</sub>	k <sub>2022</sub>	Сообщения
Оппозиция	3920	6487	0.46	0.85	0.9	1.86*10 <sup>7</sup>
Власть	3497	6050	0.28	0.6	0.8	2.3*10 <sup>7</sup>
Неопознанная	876	1398	0.07	-	-	3.42*10 <sup>6</sup>
Остальные	517	443	-	-	-	6*10 <sup>5</sup>



## МНК для политической поляризации общества ( $H_4$ )

- Спецификация:

$$\text{DesCom} = \beta_0 + \beta_1 \cdot \text{Year} + \beta_2 * \text{Old} + \beta_3 \cdot \text{Community} + \beta_4 \cdot \text{Interactions}$$

- Зависимая переменная (DesCom) - деструктивность сообщения.
- Независимая переменная - год (Year) (2021/2022), Community - сообщество каналов.



## Результаты тестирования гипотезы 4 ( $H_4$ )

	DesCom	DesCom	DesCom	DesCom	DesCom
Intercept	0.113*** (0.000)	0.074*** (0.000)	0.072*** (0.000)	0.073*** (0.000)	0.073*** (0.000)
Year	0.092*** (0.000)	0.063*** (0.000)	0.061*** (0.000)	0.066*** (0.000)	0.054*** (0.000)
Old		-0.026*** (0.000)	-0.014*** (0.000)	-0.025*** (0.000)	-0.008*** (0.000)
Opposition		-0.011*** (0.000)	-0.013*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
Pro-government		0.093*** (0.000)	0.098*** (0.000)	0.084*** (0.000)	0.084*** (0.000)
Unknown		0.114*** (0.000)	0.135*** (0.000)	0.132*** (0.000)	0.132*** (0.000)
Year*Opposition				-0.015*** (0.000)	-0.026*** (0.000)
Year*Pro-government				0.009*** (0.000)	0.027*** (0.000)
Year*Unknown				-0.024*** (0.000)	0.011*** (0.000)
Old*Opposition			-0.003*** (0.000)		0.013*** (0.000)
Old*Pro-government			-0.014*** (0.000)		-0.027*** (0.000)
Old*Unknown			-0.004*** (0.000)		-0.048*** (0.000)
N	$2*10^7$	$2*10^7$	$2*10^7$	$2*10^7$	$2*10^7$



## Интерпретация результатов тестирования гипотезы 5 (H<sub>5</sub>)

- Группа, которая связана с действующей властью выражает большую степень деструктивности, чем оппозиционная.
- Оппозиционная группа стала больше использовать деструктивную коммуникацию в 2022 году, но в два раза меньше, чем другие.
- Неопределенная группа является самым деструктивными сообществом.
- Если канал был активным в 2021 году, то он в меньшей степени деструктивный.
- Коэффициент при Year остается стабильным на во всех моделях, что может говорить о том, что политическая поляризация не является источником увеличением деструктивности.



## Выводы

- Деструктивная политическая коммуникация в русскоязычном политическом сегменте Telegram усилилась вместе с ростом его аудитории и другими общественно значимыми событиями.
- Деструктивное общение в русскоязычном политическом сегменте Telegram претерпело существенное качественное изменение в 2022 году.
- Основной вклад в развитие деструктивной коммуникации оказали темы, а не сообщения по все возможным вопросам.
- На основе сетевых данных мы увидели, что основными генераторами деструктивности являются провластные Telegram-каналы, но при этом не политическая поляризация внутри Российского общества не объясняет ее увеличение в 2022 году.



## Ограничения

- Возможность удаления как Telegram-канала, так и их сообщений.
- Несовершенство модели для измерения токсичности. Контекст BERT равен 512 токенов.
- Тегирование данных самостоятельно, без привлечения внешних аннотаторов.
- Использование только части данных для обучения тематических моделей, а также использование четких множеств для BERTopic.
- Мы сосредоточились только на текстовых данных и не рассматривали видео-, аудио- или фотоматериалы.



Спасибо за внимание