

# АВТОМАТИЧЕСКАЯ ДЕТЕКЦИЯ ТЕМАТИЧЕСКОГО ЗАРАЖЕНИЯ В TELEGRAM: ГИБРИДНЫЙ ПОДХОД С ИСПОЛЬЗОВАНИЕМ LLM И CATBOOST<sup>1</sup>

Третьяков Д.А., Савельев А.О., Карпова А.Ю.

Национальный исследовательский Томский политехнический университет, Томск, Россия  
dat32@tpu.ru, sava@tpu.ru, belts@tpu.ru

*Аннотация.* В статье представлен гибридный подход для автоматической детекции тематического заражения в Telegram, объединяющий методы семантического анализа, машинного обучения (CatBoost) и графового моделирования. Предложена методика, включающая отбор пар сообщений с помощью приближенного поиска ближайших соседей (ANN), инженерии признаков, разметку данных с использованием больших языковых моделей (LLM) и классификацию пар сообщений. Методика была апробирована на корпусе из 10,7 млн сообщений, опубликованных в более чем 500 Telegram-каналах за период с 2021 по 2024 год. Результаты демонстрируют высокую точность модели ( $ROC\ AUC = 0.8872$ ) и позволяют выявить ключевые закономерности распространения тематического заражения, включая доминирование тем, связанных с общественно-политической повесткой.

*Ключевые слова:* тематическое заражение, Telegram, машинное обучение, большие языковые модели, графовый анализ.

## Введение

Изучение социального заражения как отдельного направления исследований появилось в XXI веке. На сегодняшний день не существует консенсуса в определении социального заражения. Мы опираемся на определение данное в социологическом словаре. «Социальное заражение – contagion social – по Г. Блумеру – распространение идей, мнений, эмоциональных состояний на общественное сознание» [1].

Социальное заражение относится к намерениям и поведению человека, сформированным и находящимся под влиянием наблюдений и социализации личности [2]. То есть, аттитюды и намерения имеют значительную связь, намерения в значительной степени зависят от влияния социального заражения [3]. Социальное заражение связано со сложными когнитивными процессами, в которых восприятие, принятие, распространение идей зависит от доверия, социальных норм, намерений, мотивов, цели, а также склонности человека к подтверждению существующих убеждений (когнитивным искажениям).

Цифровая среда, включающая социальные сети и мессенджеры, такие как Telegram, коренным образом трансформировала ландшафт социального заражения, значительно ускорив и масштабировав его распространение. Онлайн-пространство обладает рядом уникальных особенностей, которые влияют на динамику и механизмы этого феномена:

- Увеличенная скорость и масштаб распространения информации в силу отсутствия традиционных географических, социальных и иных барьеров. Новость, мем, идея или эмоциональный призыв могут охватить миллионы пользователей за считанные часы или даже минуты [4].
- Сетевые каскады и «вирусное» распространение, обеспечиваемое структурой социальных графов пользователей и сообществ наряду с простыми механизмами информационного обмена (share и forward), встроенными в социальные платформы. Рекомендательные алгоритмы, обеспечивающие охват и вовлеченность аудитории в свою очередь могут способствовать как «вирусности», так и «забвению» контента.
- Компьютерно-опосредованная коммуникация (Computer-Mediated Communication, CMC) формирует уникальные типы социальных связей и онлайн-сообществ (например, VK-сообщества, Telegram-каналы), которые характеризуются высокой плотностью внутренних связей и гомофилией что также способствует ускоренному распространению идей и эмоций, а также формированию «эхо-камер» или «фильтрационных пузырей» [5].
- Низкая «стоимость» коммуникаций и распространения контента, достигаемая за счет технических и функциональных возможностей социальных медиа.
- Асинхронность и возможность отложенной реакции, обеспечивающие пользователя временем для обдумывания и формирования ответа, также способны изменять динамику эмоционального заражения.

---

<sup>1</sup> Публикация подготовлена в рамках проекта РНФ № 25-28-01153 «Моделирование механизмов социального заражения в процессе онлайн-радикализации».

- Анонимность и псевдонимность. Некоторые платформы позволяют пользователям действовать анонимно или под псевдонимами, что сказывается на их поведении, снижая социальные ограничения и, потенциально, способствуя распространению более смелых или даже деструктивных идей и эмоций.

Эти особенности цифровой среды создают уникальный контекст для изучения тематического заражения, требующий адаптации существующих моделей и разработки новых подходов, учитывающих специфику онлайн-платформ, таких как Telegram, с их каналами, возможностями пересылки сообщений и относительно открытым доступом к контенту.

Таким образом, целью работы является разработка механизмов для автоматической детекции признаков социального заражения в социальных медиа, а именно – признаков тематического заражения. Социальное заражение в онлайн-коммуникации в самом общем виде мы рассматриваем как «тенденцию пользователей социальных сетей совершать такие действия в сети как повторное размещение, обмен информацией или принимать новое поведение после воздействия аналогичной информации или поведения соответственно» [6]. Термин «тематическое заражение» ещё не имеет устоявшегося определения. Для цели исследования мы определяем тематическое заражение как процесс распространения контента в виде выражения идеи, рассуждения или размышления, ограниченного рамками тематического фрейма, посредством которого пользователи социальных сетей передают основную мысль, формируют установки и конструируют индивидуальные и коллективные представления. В результате пользователи вовлекаются в процесс социального заражения, а их поведенческие решения и действия находятся в прямой зависимости от воздействия управляющих субъектов.

## **1. Социальное заражение в процессе сетевого взаимодействия (онлайн-коммуникаций) как отдельное направление исследований**

Ключом к пониманию распространения социального заражения является структура сетей, лежащих в основе сильных и слабых связей, которые складываются в процессе сетевого взаимодействия. Концепция «сила слабых связей» Марка Грановеттера [7], с 70-х гг. XX века стала основой в исследованиях онлайн-коммуникаций.

Социальное заражение имеет корни в социальной структуре, побуждая людей в аналогичных социальных ситуациях оценивать преимущества и риски принятия решений аналогичных ситуациях [8]. Когда люди взаимодействуют с различными социальными факторами, они склонны усваивать неявные нормы выбора, которые становятся фундаментальной основой для их будущих решений [2]. Соответственно, принятие решений будет происходить либо невольно как результат воздействия «контактного окружения» человека, либо добровольно, осознанно.

В одном из масштабных исследований по изучению крупномасштабного процесса заражения – пример социального процесса глобального распространения социального заражения в Facebook<sup>2</sup>. Авторы обнаружили несколько интересных и новых признаков, которые ранее были неизвестны. Суть их заключается в следующем: (1) вероятность социального заражения жестко контролируется количеством связей в «контактном окружении» человека, а не фактическим размером этого окружения; (2) если такое «структурное разнообразие» контролируется, то размер «контактного окружения» становится негативным предиктором социального заражения; (3) данные такого масштаба как изучение глобального заражения в сети Facebook позволили идентифицировать «структурные сигналы», которые в меньших масштабах изучения оказались не замечены, но, они играют критически важную роль в прогнозировании результатов социальных процессов; (4) необходимо переосмыслить механику, лежащую в основе функционирования процесса социального заражения, вместо того, чтобы рассматривать количество в окружении человека, как решающий параметр, надо рассматривать количество контекстов, которые представляет окружение, как движущий механизм социального заражения; (5) роль «структурного разнообразия» в процессе социального заражения указывает на интересные направления дальнейших исследований для математического моделирования процесса; (6) глобальные феномены возникающие в результате процесса заражения в сетях (где поведение в узле имеет нетривиальную зависимость от полного набора поведения в соседних узлах) можно включить в имитационные модели для изучения базовой вероятности заражения, например, на количестве связанных компонентов, образованных затронутых соседями узла. Тогда возможно найти понимание

---

<sup>2</sup> Деятельность организации запрещена на территории РФ

того, как меняются глобальные свойства процессов заражения; (7) предлагается потенциальный пересмотр основных теорий о роли, которую сети играют в социальной и экономической сферах [9].

Заражение положительными и отрицательными эмоциями, по мнению некоторых исследователей имеет схожие черты, что в принципе соответствует некоторым данным, которые были изучены на материалах оценки поведения в офлайн [10] и даже экспериментам, в которых изучалось заражение с использованием нейровизуализации [11].

Есть и другое мнение, основанное на результатах экспериментов, в которых доказано, что положительные эмоции могут играть ключевую роль в эмоциональном заражении как в онлайн, так и в офлайн [12-14]. Такой вывод идет вразрез со сложившимися в предыдущем столетии представлениями о том, что люди сильнее реагируют на отрицательные аттитюды [15].

Цифровые медиа вносят большой вклад в формирование и распространение социальных движений в Интернете. Но, до сих пор остается открытым вопрос о том, в какой степени эти движения преобразуются в конкретные действия в офлайне. За прошедшее десятилетие эксперты неоднократно отмечали, что радикальные и ультрарадикальные движения в социальных сетях чаще стали позиционировать свою идеологию мимикрируя под менее радикальные движения. Например, тактическое новаторство ультраправых заключается в том, что уличный активизм был в значительной степени заменен онлайн-активизмом, в котором прямое насилие становится менее вероятным. Некоторые ученые ссылаются на растущие сообщества, называющие себя Alt-light (правое политическое движение, члены которого считают себя отделенными как от основного консерватизма, так и от крайне правой, белой националистической точки зрения) как более интеллектуальное движение, которое предпочитает долгосрочную, демократическую медиа-ориентированную активность против насильственной уличной активности и терроризма [16].

Перевод онлайн-активизма в коллективные действия за пределами социальных сетей на настоящий момент времени, имеет подтверждение ограниченности применения. Большое количество социальных связей пользователей и высокая частота социальных взаимодействий приводят к более частым социальным онлайн-движениям, росту онлайн-активизма [17]. Например, такие социальные движения как «Арабская весна» или Black Lives Matter не смогли бы получить такого широкого, мирового распространения, узнаваемости и отклика у широкой общественности, без цифровых медиа. По результатам многочисленных исследований, эксперты отмечают, что движущей силой распространения социальных движений и онлайн-активизма является обмен эмоциями между пользователями, который вызван негативным откликом, т.е. вызывает гнев [19-21].

Некоторые исследователи высказывают предположение, что если модерации будет подвергаться негативный контент, то, пользователи значительно меньше будут проявлять отрицательных эмоций. Кроме того, с учетом доказательств того, что компании цифровых медиа желают максимизировать эмоции пользователей, некоторые исследователи предполагают, что алгоритмы цифровых медиа особенно продвигают контент с положительными эмоциями [22].

В настоящее время, используя вычислительные методы социальных наук для анализа крупномасштабных данных социальных сетей, поведение людей в онлайн-коммуникациях, безусловно можно изучать с большей статистической мощностью и эмпирической обоснованностью. Однако, как утверждают исследователи «текущие стандарты проверки значимости нулевых гипотез и корреляционной статистики, по-видимому, плохо подходят для заметно шумных, многомерных наборов данных социальных сетей» [23].

Всё вышеизложенное актуализирует задачу разработки соответствующего научно-технического инструментария для автоматизации процессов изучения механизмов социального заражения в онлайн-среде.

## **2. Методика автоматической детекции признаков тематического заражения**

Для детекции признаков тематического заражения в Telegram был разработан гибридный подход, объединяющий методы семантического анализа, машинного обучения и графового моделирования. Процесс включает несколько основных этапов, направленных на выявление и кластеризацию сообщений, связанных с одним новостным событием (далее – ивентом):

1. Отбор пар-кандидатов. Из всего массива сообщений (более 10 млн) с помощью приблизительного поиска ближайших соседей (Approximate Nearest Neighbor, ANN) на основе Faiss [24] отбираются пары сообщений, потенциально относящиеся к одному событию. Для этого используются предварительно рассчитанные семантические эмбединги текстов, полученные с помощью моделей семейства E5 [25]. Пороги косинусного сходства варьируются от 0,96 до 1 для обеспечения баланса между точностью и полнотой.

2. Инженерия признаков. Для каждой пары сообщений генерируется более 100 признаков, включая временные характеристики (разница во времени публикации), пересечение контента (URL, хэштеги, именованные сущности), наличие общих новостных ссылок, метаданные (просмотры, репосты) и признаки информативности текстов. Инженерия признаков позволяет количественно оценить сходство и различия между сообщениями.

3. Разметка данных с использованием больших языковых моделей (далее – LLM). Для обучения модели CatBoost применяется итеративный процесс разметки пар сообщений с помощью LLM. Первоначально оценивалась информативность текстов, затем определялась их принадлежность к одному событию. Для спорных случаев использован арбитраж с уточнением меток для обеспечения качества обучающей выборки.

4. Обучение модели CatBoost. На основе признаков и LLM-меток обучена модель CatBoost для классификации пар сообщений. Подбор гиперпараметров выполнен с помощью Optuna – фреймворка для автоматизированного поиска гиперпараметров [26], а оценка качества – на отложенной тестовой выборке. Модель демонстрирует высокую точность (ROC AUC = 0.8872).

5. Формирование ивентов. Результаты предсказания модели CatBoost использованы для построения графа связей между сообщениями. Компоненты связности в графе определяются как ивенты – кластеры сообщений, связанных общей темой. Данные загружены в графовую СУБД для дальнейшего анализа и визуализации.

6. Анализ и интерпретация. Было выполнено исследование характеристик ивентов (масштаб, длительность, популярность) и их тематической направленности с использованием инструментов графовой аналитики.

### **3. Результаты детекции признаков тематического заражения в публичных Telegram-каналах**

После обучения и оценки модели CatBoost, она была применена ко всему набору из более чем 108 миллионов пар-кандидатов сообщений для предсказания вероятности их принадлежности к одному новостному событию. На основе этих предсказаний и выбранного порога классификации (в данном детальном анализе используется порог 0.9461, который был выбран как один из оптимальных в ходе экспериментов для достижения баланса между точностью и полнотой) были построены графы связей между сообщениями. Компоненты связности в этих графах были определены как "ивенты" – группы сообщений, объединенные общей темой или новостным поводом, что и представляет собой результат детекции тематического заражения. Данные об ивентах и участвующих в них сообщениях и каналах были загружены в графовую базу данных для последующего анализа.

#### **Обзор и общая характеристика выявленных событий (ивентов)**

Анализ масштаба выявленных ивентов проводился по трем основным показателям: количество сообщений в ивенте, количество уникальных каналов-участников и длительность ивента. По анализируемой выборке среднее количество сообщений в ивенте составило 4,23; медианное значение сообщений – 2; наименьшее количество – 1 и наибольшее – 1 009 377 при стандартном отклонении в 1 053,12. Из этого можно заключить, что большинство выявленных ивентов «состоят» из небольшого количества сообщений. Однако наличие ивентов с максимальным количеством сообщений, достигающим более миллиона, и высокое стандартное отклонение указывают на сильно скошенное вправо распределение («длинный хвост»). Это означает, что существует небольшое количество очень крупных, вирусных событий, которые охватывают значительное число сообщений, в то время как основная масса ивентов представляет собой мелкие тематические кластеры.

Среднее количество каналов, участвующих в ивенте, равно 2,61; медианное значение каналов – 2; наименьшее – 1; наибольшее – 570 при стандартном отклонении в 3,41. Аналогично количеству сообщений, большинство ивентов затрагивают небольшое число каналов. Максимальное значение в 570 каналов для одного ивента свидетельствует о способности некоторых тем охватывать значительную часть активной Telegram-экосистемы (в рамках исследуемых 596 каналов). На рисунке 1 приведен график распределения ивентов по количеству участвующих в них Telegram-каналов.

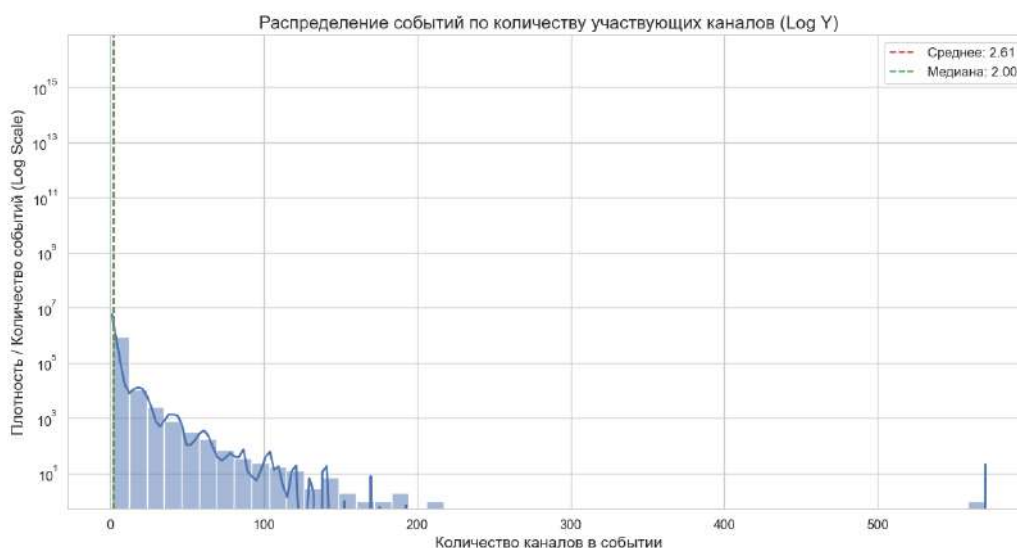


Рис. 1. Распределение ивентов по количеству участвующих каналов (Логарифмическая шкала по оси Y)

Средняя длительность ивента в часах составила 241,35; медианная длительность – 0,75; наименьшая – 0; наибольшая – 35 058 при стандартном отклонении в 1 766,86. Медианная длительность ивента менее часа указывает на то, что многие тематические всплески являются кратковременными. Однако максимальная длительность в почти 4 года и высокое стандартное отклонение свидетельствуют о наличии очень долгоживущих тем или о способности модели объединять сильно разнесенные по времени, но семантически связанные сообщения в один ивент. Нулевая минимальная длительность может означать наличие ивенты, состоящих только из сообщений, опубликованных практически одновременно.

#### Популярность и временная динамика ивентов

Популярность ивентов, вычисляемая на основе количества просмотров соответствующих сообщений, демонстрирует схожий характер с метриками общих характеристик: среднее значение – 78 533,3; медианное значение – 33 194,57; наименьшее – 0; наибольшее – 4 341 822. Существуют ивенты, сообщения в которых в среднем набирают миллионы просмотров, в то время как медианное значение значительно ниже. Это указывает на то, что лишь немногие темы вызывают значительный общественный резонанс.

Анализ количества и характеристик ивентов по месяцам позволяет выявить периоды повышенной информационной активности и возможные изменения в характере тематического заражения. На рисунке 2 и 3 приведены соответственно графики общего количества выявленных ивентов и среднего количества каналов, участвующих в ивенте, по месяцам.

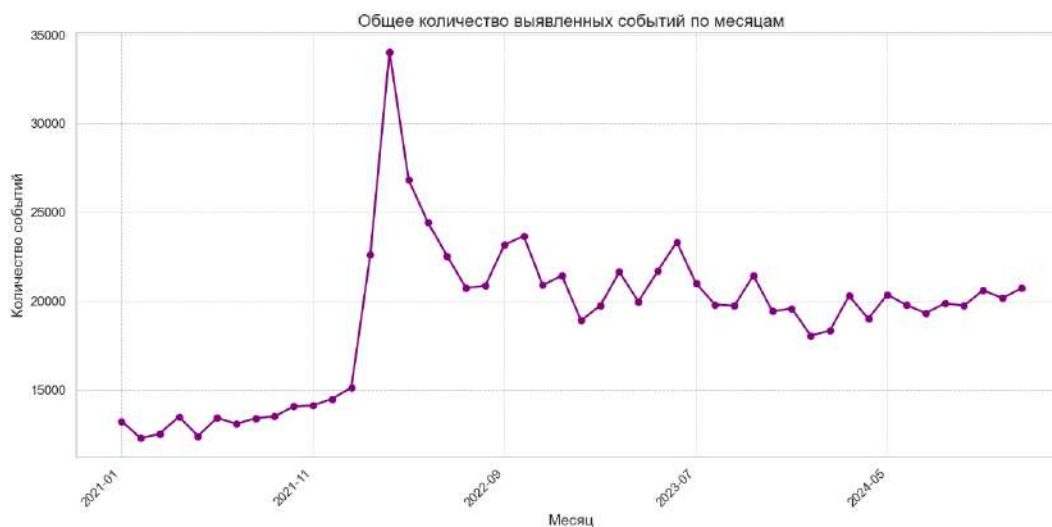


Рис. 2. Общее количество выявленных ивентов по месяцам

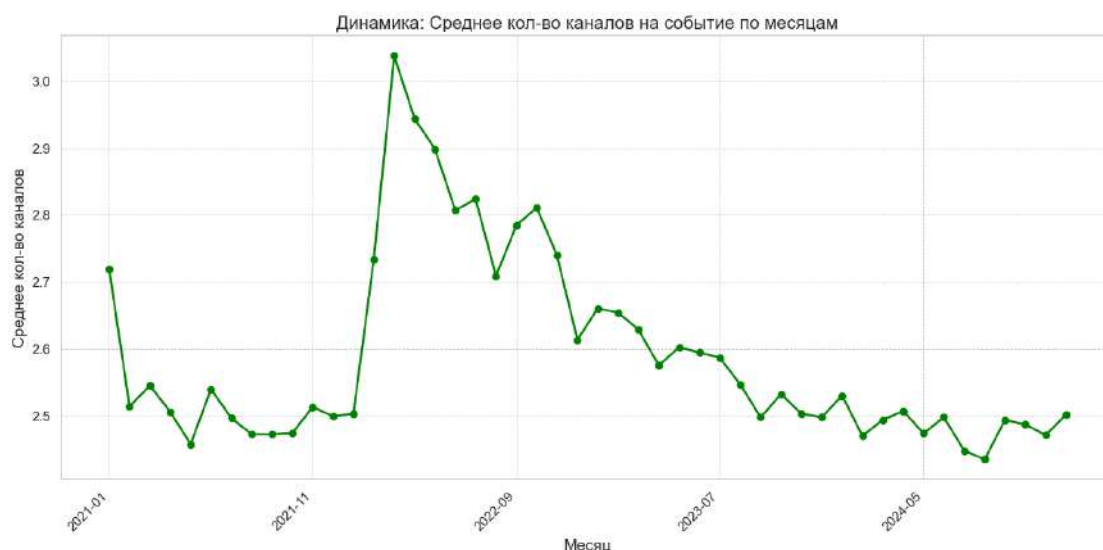


Рис. 3. Среднее количество каналов, участвующих в ивенте, по месяцам

Отметим, что наблюдается значительный рост количества ивентов, начиная с начала 2022 года. Средняя популярность ивентов может коррелировать с периодами высокой общественной напряженности или интереса к определенным темам.

#### Именованные сущности, связанные с ивентами

Для понимания тематической направленности выявленных ивентов был проведен анализ наиболее часто встречающихся именованных сущностей (NER), агрегированных на уровне ивентов. В таблице 1 приведены 5 наиболее часто упоминаемых именованных сущностей по категориям. На рисунке 4 приведен график общего количества 20 наиболее часто упоминаемых именованных сущностей-персон.

Таблица 1. Топ-5 наиболее часто встречающихся в ивентах именованных сущностей

Тип именованной сущности	1	2	3	4	5
Персоны	Владимир Путин	Владимир Зеленский	Джо Байден	Дональд Трамп	Дмитрий Песков
Локации	Российская Федерация	Украина	США	Москва	Киев
Организации	ВСУ	СМИ	Государственная дума РФ	Министерство обороны РФ	НАТО

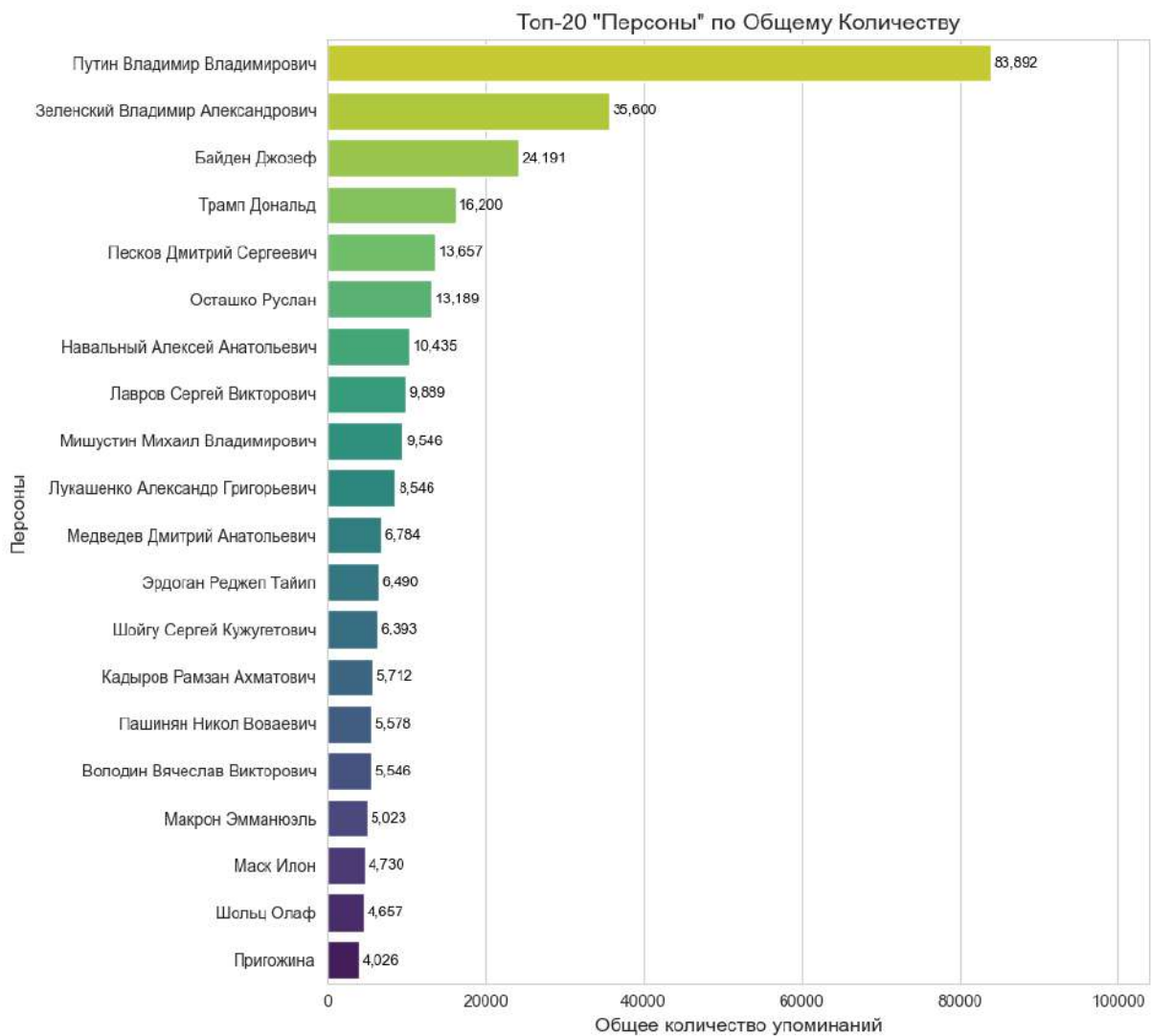


Рис. 4. 20 наиболее часто упоминаемых в выявленных ивентах персон

Также был выполнен анализ согласованности NER в ивентах (на основе выборки 200 ивентов):

- Persons: Средняя согласованность = 0.881, Медианная = 1.000.
- Locations: Средняя согласованность = 0.935, Медианная = 1.000.
- Organizations: Средняя согласованность = 0.858, Медианная = 1.000.

Высокие значения медианной согласованности для всех типов именованных сущностей указывают на то, что сообщения внутри одного автоматически выявленного ивента действительно сильно сфокусированы вокруг одних и тех же персон, локаций и организаций, что подтверждает качество детекции тематической связности событий.

#### Значимые каналы в процессе тематического заражения

Для выявления наиболее активных участников процессов тематического заражения и значимых инициаторов распространения информации был выполнен анализ участия Telegram-каналов в выявленных ивентах:

1. Наиболее активные каналы (по общему участию в ивентах): swodki (121,630 ивентов), SolovievLive (89,989 ивентов), vestiru24 (69,971 ивентов), vestiru (69,970 ивентов), bbbreaking (61,375 ивентов).

2. Каналы с наибольшим вкладом сообщений в ивенты: swodki (214,499 сообщ.), SolovievLive (158,894 сообщ.), vestiru24 (115,003 сообщ.), vestiru (115,002 сообщ.), bbbreaking (101,572 сообщ.).

3. Основные инициаторы ивентов (каналы, опубликовавшие первое сообщение в ивенте): vestiru (45,210 ивентов), vestiru24 (45,210 ивентов), bbbreaking (41,536 ивентов), swodki (30,103 ивентов), SolovievLive (29,755 ивентов). Отмечено, что топ-20 каналов-инициаторов ответственны за ~41% всех выявленных ивентов.

4. Скорость реакции каналов: Медианное время реакции (время от первого сообщения в ивенте до первого сообщения канала в этом же ивенте) составило 0.01 часа. Выявлены как очень "быстрые"

каналы (ARTolmachev, ombudsmanrf, oldhellboy – медианное время реакции 0.00 ч), так и "медленные" (например, StudyBrand369 – 46 часов).

#### 4. Обсуждение результатов и ограничения предложенной методики

Анализ выявленных тематических кластеров (ивентов) в публичных Telegram-каналах демонстрирует высокую степень тематической гомогенности дискурса, с доминированием тем, связанных со специальной военной операцией. Полученные данные свидетельствуют, что данная тематика формирует ядро общественно-политической повестки, предлагаемой аудитории. Высокий уровень вовлеченности пользователей, выраженный в показателях охвата ивентами, может интерпретироваться как индикатор соответствующего интереса целевой аудитории, либо как проявление феномена думскроллинга (doomscrolling) – компульсивного потребления тревожного новостного контента, что требует дальнейшей верификации в рамках специального исследования.

Особого внимания заслуживает выраженная персонификация тематического контента, что соответствует концепции «персонифицирующей акцентуации» Бурдые. В исследуемом дискурсе наблюдается три основных вектора реализации данного механизма:

- концентрация внимания на конкретной личности;
- акцентирование функциональной роли персоналии;
- фокусировка на политических сообщениях, транслируемых персоной.

Примечательно, что аналогичный механизм персонификации применяется и к институциональным акторам (например, НАТО), где акцент смещается на функциональную роль организации в публичном дискурсе. Особенно рельефно данный механизм проявляется в случаях медиатизации конфликта персонифицированных политических интересов. Таким образом, персонифицирующая акцентуация выступает ключевым инструментом фреймирования социальной реальности в исследуемом информационном пространстве.

В качестве ограничений предлагаемой методики, автоматизированной детекции признаков тематического заражения можно выделить:

- зависимость от качества эмбедингов – точность семантического поиска зависит от выбранных моделей, которые могут недостаточно хорошо отражать контекст коротких или неформальных сообщений;
- качество LLM-разметки зависит, с одной стороны, от промтов и, с другой стороны – от выбранной LLM;
- жесткий порог классификации CatBoost (0.9461) может приводить к избыточной фрагментации или, наоборот, объединению несвязанных событий;
- методика учитывает исключительно текстовый контент, оставляя без внимания мультимедийные вложения.

Частично указанные ограничения могут быть преодолены в дальнейшем с помощью использования мультимодальных эмбедингов и оптимизации LLM-разметки на базе few-shot обучения и доработанных промтов.

#### 5. Заключение

В данной работе предложен гибридный подход для автоматической детекции тематического заражения в Telegram, сочетающий методы семантического поиска (ANN на основе эмбедингов), разметку с использованием больших языковых моделей (LLM) и классификацию пар сообщений с помощью CatBoost. Подход был апробирован на корпусе из 10,7 млн. сообщений, опубликованных за период с 2021 по 2024 гг. в более чем 500 публичных Telegram-каналах. Несмотря на значительную вычислительную ресурсоёмкость и наличие указанных в предыдущем разделе ограничений, разработанный инструментарий позволяет выявлять множества сообщений в социальных медиа, относящихся к одному событию (информационному поводу). Это обеспечивает возможность выявления и последующего анализа примеров цифрового социального заражения для изучения соответствующих механизмов, в том числе при решении задач моделирования распространения информации, идей, убеждений и моделей поведения.

#### Литература

1. *Кравченко С.А.* Социологический энциклопедический словарь русско-английский словарь: Более 10 000 единиц. – М.: ООО «Издательство АСТ», 2004. – 511 с.

2. *Chaouali W., El Hedhli K.* Toward a contagion-based model of mobile banking adoption // *International Journal of Bank Marketing*. – 2018. – Vol. 37(1). – P. 69–96.
3. *Al-Emran M., Arpacı I., Salloum S.A.* An empirical examination of continuous intention to use m-learning: An integrated model // *Education and Information Technologies*. – 2020. – Vol. 25. – P. 2899–2918.
4. *Pariser E.* *The Filter Bubble: What the Internet is Hiding from You*. – NY.: Penguin Press, 2012. – 304 p.
5. *Zhuravskaya E., Petrova M., Enikolopov R.* Political Effects of the Internet and Social Media // *Annual Review of Economics*. – 2020. – Vol. 12. – P. 415–438.
6. *Bartal A., Pliskin N., Tsur O.* Local/ Global contagion of viral/non-viral information: Analysis of contagion spread in online social networks // *PLoS ONE*. – 2020. – Vol. 15(4).
7. *Granovetter M.S.* 1973. The Strength of Weak Ties // *The American Journal of Sociology*. – 1973. – Vol. 78 (6). – P. 1360–1380.
8. *Kiogothe M.W.* Factors influencing adoption of mobile banking in Kenya: a case of commercial banks' customers in Nairobi County. <https://su-plus.strathmore.edu/handle/11071/6163> (дата обращения 30.05.2025).
9. *Ugander J., Backstrom L., Marlow C., Kleinberg J.* Structural diversity in social contagion // *Proceedings of the National Academy of Sciences of the United States of America*. – 2012. – Vol. 109(16). – P. 5962–5966.
10. *Barsade S.G.* The ripple effect: Emotional contagion and its influence on group behavior // *Administrative Science Quarterly*. – 2002. – Vol. 47. – P. 644–675.
11. *Willroth E.C., Koban L., Hilimire M.R.* Social information influences emotional experience and late positive potential response to affective pictures // *Emotion*. – 2017. – Vol. 17. – P. 572–576.
12. *Coviello L., Sohn Y., Kramer A.D., Marlow C., Franceschetti M., Christakis N.A., Fowler J.H.* Detecting emotional contagion in massive social networks // *PLoS One*. – 2014. – Vol. 9.
13. *Bösch K., Müller O., Schneider J.* Emotional contagion through online newspapers // *Research Papers*. – 2018. – Vol. 171.
14. *Bhullar N.* Self-ratings of love and fear on emotional contagion scale depend on the environmental context of rating // *Current Research in Social Psychology*. – 2012. – Vol. 19.
15. *Cacioppo J.T., Gardner W.L., Berntson G.G.* Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space // *Personality and Social Psychology Review*. – 1997. – Vol. 1(1). – P. 3–25.
16. *Ravndal J.A.* Right-wing terrorism and violence in Western Europe: Introducing the RTV dataset // *Perspectives on Terrorism*. – 2016. – Vol. 10(3).
17. *Tufekci Z.* *Twitter and tear gas: the power and fragility of networked protest*. Yale University Press; 2017. – 360 p.
18. *Anduiza E., Cantijoch M., Gallego A.* Political participation and the internet // *Information, Communication & Society*. – 2009. – Vol. 12. – P. 860–878.
19. *Van Zomeren M., Leach C.W., Spears R.* Protesters as “passionate economists”: A dynamic dual pathway model of approach coping with collective disadvantage // *Personality and Social Psychology Review*. – 2012. – Vol. 16. – P. 180–199.
20. *Crockett M.J.* Moral outrage in the digital age // *Nature Human Behaviour*. – 2017. – Vol. 1. – P. 769–771.
21. *Castells M.* *Networks of outrage and hope: Social movements in the internet age*. Polity Press, 2012. – 345 p.
22. *Lee S.Y., Hansen S.S., Lee J.K.* What makes us click like on Facebook? Examining psychological, technological, and motivational factors on virtual endorsement // *Computers Communications*. – 2016. – Vol. 73. – P. 332–341.
23. *Burton J.W., Cruz N., Hahn U.* Reconsidering evidence of moral contagion in online social networks // *Nature Human Behaviour*. – 2021. – Vol. 5. – P. 1629–1635.
24. Faiss. <https://github.com/facebookresearch/faiss> (дата обращения 30.05.2025).
25. Multilingual-E5-large-instruct. <https://huggingface.co/intfloat/multilingual-e5-large-instruct> (дата обращения 30.05.2025).
26. Optuna. <https://optuna.org> (дата обращения 30.05.2025).