
ИНТЕРНЕТ-ИССЛЕДОВАНИЯ



DOI: 10.19181/4m.2025.34.2.4

EDN: IUUUME

ПЕРСПЕКТИВЫ МЕДИА-МОНИТОРИНГА В ИССЛЕДОВАНИЯХ ОБЩЕСТВЕННОГО МНЕНИЯ (НА ПРИМЕРЕ ДОВЕРИЯ ПРЕЗИДЕНТУ)

Анкудинов Иван Андреевич

Национальный исследовательский университет

«Высшая школа экономики»,

Москва, Россия

Эл. почта: iankudinov@hse.ru

ORCID: 0000-0002-7001-4163

Для цитирования: Анкудинов И. А. Перспективы медиа-мониторинга в исследованиях общественного мнения (на примере доверия Президенту) // Социология: методология, методы, математическое моделирование (Социология: 4М). 2025. № 61. С. 165-203. DOI: 10.19181/4m.2025.34.2.4. EDN: IUUUME.

Изменчивые политические настроения россиян — постоянный предмет интереса социологических фондов. С развитием интернета привычные анкетные исследования стали дополняться онлайн-опросами и, несмотря на некоторый скепсис, «майнингом» социальных сетей. В настоящей статье предпринимается попытка скорректировать стихийную интернет-выборку так, чтобы приблизить ее оценки к репрезентативным омнибусам. Мы используем показатели доверия Президенту РФ в сети и в опросах с целью ответить на вопросы: насколько далеки друг от друга эти источники? Можно ли сблизить их оценки за счет постстратификационных весов? Наружено ли отношение между ними началом СВО? Является ли это отношение (несмотря на какую-то абсолютную разницу) стабильным во времени, т. е. представляется ли возможным перейти от одного индикатора к другому за счет линейных преобразований? Выводы, к которым мы приходим по итогам

статистического анализа, таковы: (1) интернет-замеры, основанные на усредненной словарной тональности, оценивают доверие Президенту на 20 п. п. ниже, чем результаты опросов (2). Коррекция выборки пользователей на веса из переписи населения не позволяет сблизить сетевые и опросные оценки (3). Уточнение оценок по конкретным площадкам и поллстера姆 лишь подтверждает обнаруженную разницу. Для некоторых платформ эта разница более существенна, что вполне объяснимо спецификой их аудитории (4). Факт начала СВО неоднозначным образом отразился на различиях между оценками. Тональность по «Медиалогии» фиксирует увеличение разрыва, в то время как Dostoevsky почти не видит разницы «до» и «после» (5). Прийти к однозначному заключению о темпоральной зависимости онлайн-и онлайн-оценок нельзя. Взвешивание интернет-выборки дает некоторые основания считать ряды «сети» и «поля» коинтегрированными, но эта связь проявляется не во всех тестах.

Ключевые слова: поллстеры, социальные сети, взвешивание, анализ настроений, анализ текстов, временные ряды, президентский рейтинг, политическое доверие

Введение

26 мая 2023 года на страницах американской The New York Times вышла статья, озаглавленная «Russian Public Appears to Be Souring on War Casualties, Analysis Shows» [«Анализ показывает, что российская общественность, похоже, разочарована потерями на войне»] [1]. Анализ, на который ссылалась газета, был проведен компанией FilterLabs AI на основе публикаций в социальных сетях и интернет-форумах. Он показывал, что со временем россияне стали более критичны в своих суждениях об СВО даже с поправкой на майские праздники, иронию, ботов и другой текстовый «шум». Сама статья не вызвала большого ажиотажа, поскольку ее выводы лишь подкрепляли сложившийся в западной академии консенсус [см., например, [2]], но оказалась интересна методологически: она представила еще один способ «альтерна-

тивной социологии» для контекста, в котором традиционная (по тем или иным причинам) подозревается в искажениях.

NYT и FilterLabs привели довольно типичную мотивацию для использования сетевых данных «поверх» опросных: в условиях кризиса респонденты подвержены давлению и самоцензуре, а потому склонны фальсифицировать предпочтения и завышать долю социально одобряемых ответов. Добавим, что именно в таких условиях общественное мнение играет ключевую роль, поскольку служит редким сигналом к коллективному действию [3]. С другой стороны, использование веб-данных также чревато потерями в репрезентативности, поскольку активные пользователи сети, как правило, составляют лишь долю от общего населения [4; 5]. Это противоречие, положенное на специфический контекст современной России, позволяет нам задаться вопросом: какова связь между этими двумя (по-своему смещёнными) способами замера общественных настроений?

Формально говоря, нас интересует, в какой степени интернет-активность россиян на политическую тематику коррелирует с результатами репрезентативных опросов по той же теме. В качестве референтного показателя, предположительно, наиболее важного для персоналистских систем, мы рассматриваем президентский рейтинг; в РФ, он регулярно измеряется крупными поллстерами — ВЦИОМ, ФОМ и «Левада-центром»¹. В то же время фигуру Президента постоянно обсуждают в сети, причем информационный контекст там кажется более насыщенным и актуальным. Это дает основания предположить, что разброс оценок там также будет больше, но существенные переломы в настроениях граждан должны быть отражены и сетью, и поллстерами, то есть их можно использовать для своеобразной кроссвалидации.

¹ АНО «Левада-Центр» внесена в реестр некоммерческих организаций, выполняющих функции иностранного агента.

Для проверки этой гипотезы мы используем результаты тематических опросов за полгода до и полгода после 24.02.2022, а также агрегированную тональность постов в Рунете за тот же период. Выбор такого временного промежутка позволяет одновременно проверить, не была ли предполагаемая связь нарушена началом СВО — переломным политическим событием последних нескольких лет. Чтобы извлечь из всего потока публикаций релевантные тексты, мы прибегаем к поиску по ключевым словам и машинной разметке, а затем исследуем полученные временные ряды на коинтеграцию. Эта трехступенчатая процедура несовершенна, но позволяет дать количественную оценку сходству опросной и «сетевой» динамики. Целью статьи является, собственно говоря, получение этой оценки.

Текст статьи структурирован следующим образом. В первом разделе мы останавливаемся на особенностях сетевых и социологических измерений и обозначаем перспективы их взаимозамены. Во втором разделе представлен общий вид моделей, используемых для проверки связи «сети» и «поля». В третьей части представляются результаты оценивания этих моделей. В заключении мы обсуждаем возможные причины наблюдаемых результатов, ограничения и планы будущих разработок.

Опросная социология и ее субSTITУты

Даже самые оптимистичные обзоры показывают, что выработать лучший механизм выявления общественных предпочтений, чем репрезентативные опросы, в ближайшее время не получится [6, р. 110713]. При этом, несмотря на сравнительную точность и статистическую достоверность², исследования со стандартизованными анкетами имеют некоторые очевидные недостатки,

² То есть измеримые ошибки и возможность инференции на более широкую совокупность.

борьба с которыми до сих пор порождает методологические споры. Разнообразные смещения выборки из-за недостижимости большей части потенциальных респондентов, дороговизна полевой работы, ограничения, возникающие из-за формулировок вопросов и вариантов ответов, слабая предсказательная сила результатов — все это наносит ущерб репутации публичной социологии [7]. Вышеназванное усугубляется в закрытых политических режимах, где социально одобряемая позиция явно артикулирована, а оппозиционные взгляды в значительной мере маргинализованы. Как следствие, искажения, заложенные в дизайн опроса, накладываются на субъективные переживания респондентов, в разной мере подверженных социальному давлению; возможности по манипулированию опросными данными растут, а доверие к ним падает [8].

Конкуренцию опросной традиции в разное время составляли разные подходы. Авторы тематических сборников и обзорных статей [9; 10; 11] приводят как минимум четыре конкурирующие парадигмы:

- а) сознательный отказ от опросов. Поскольку общественного мнения, по утверждению одного известного социолога, не существует, попытки его измерения создают ложное ощущение вовлеченности публики в происходящие (в частности, политические) процессы. Ориентиром, базовой сравнительной категорией для исследователей должен быть последний факт социального действия: например, результаты выборов или массовость последовавших за ними протестов. Этот подход ограничивает нас в возможности отслеживать динамику настроений, но в то же время избавляет от нужды усреднять противоречивые, нередко полярные даже внутри одной анкеты мнения;
- б) отказ от количественных опросов в пользу качественных интервью. Возможность напрямую беседовать с

гражданами, записывая их полнотекстовые ответы и невербальные реакции, регулируя «глубину» дискуссии и акцентируя внимание на интересующих нарративах, привлекает своей относительной простотой. Эта простота, однако, достигается за счет отказа от распространения результатов за пределы опрашиваемой группы, т. е. за счет внешней валидности. В условиях какой бы то ни было цензуры (в том числе самоцензуры), выводы любого из таких интервью легко поставить под сомнение;

- в) отказ от интерпретации единичных опросов в пользу агрегаторов общественного мнения. Логика многократного усреднения и взвешивания (по респондентам с различным жизненным опытом, по разным социальным группам, по опросам, проведенным в отличающихся контекстах и пр.) позволяет считать значительное число третьих факторов рандомизированными, а потому не влияющими на наблюдаемый результат. Это рассуждение справедливо, хотя и страдает от тех же аксиоматических упреков, которые можно предъявить индивидуальным опросам; в частности, усредненный ответ на неудачно сформулированный вопрос не приближает нас к какой-то «истинной» позиции общественности, если допустить, что таковая вообще существует;
- г) переориентация на большие данные социальных медиа. С одной стороны, объемы этих данных выполняют ту же функцию «взаимопогашения» экстремальных значений, что и агрегаторы из пункта выше, а проблему переведенности в сети более молодой и активной аудитории можно обернуть в свою пользу: целью политического анализа нередко является именно вовлеченная аудитория, способная к инициативному действию. С другой стороны, невозможность проконтролировать

обстоятельства, стоящие за написанием конкретных постов, ведут к большой вариации замеров. Кроме того, не существует никакой конвенциональной методологии сэмплирования и обработки сетевых данных, из-за чего результат становится зависимым от качества векторизации текстов, фильтрации спама и последующей маркировки.

Конечно, авторы оговариваются, что названные подходы призваны дополнить, а не заменить регулярную работу поллстлеров. Напоминание о том, что общественные настроения почти никогда четко не сформулированы и чересчур дисперсны, полезно при интерпретации любых омнибусов. Предпочтение глубинных интервью (то есть, в конечном итоге, больших текстов) указывает на проблему формулировок, которая является ценой стандартизации в анкетных опросах. Консервативный метод межопросного агрегирования позволяет избежать «переобучения», скажем, на изменчивой кривой политических рейтингов, но усугубляет проблему дефицита ресурсов и вынуждает к использованию статистических *post hoc* техник (*bootstrapping*, *over-* и *undersampling*). О сетевых данных чаще вспоминают, когда речь идет о выявлении остросоциальной проблематики или о том, что в профильной литературе называют «анализом мнений» (*opinion mining*) [см. 12].

Мы полагаем, что эти аргументы можно объединить, показав, что сетевой метод является ключевым комплементом (если не субститутом) опросного. Лишний раз констатируем, что этот метод имеет несомненные упущения; к уже названному стоит добавить, что активность в не-модерируемых микроблогах может поддерживаться ботами, а социально-демографическое профилирование в сети затруднено больше, нежели в любой форме опроса [13; 14]. Если, например, американский X (бывший Twitter) снискал популярность среди разных слоев англоязычной аудитории, в России онлайн-площадки имеют ярко выра-

женные «перекосы», в том числе из-за различий в потребляемом контенте [15]. Эти диспропорции только усиливаются из-за вмешательства федерального регулятора, активизировавшего практику блокировок в 2022–2023 гг., а также из-за рисков деанонимизации [16; 17].

Но, как мы уже успели заметить, политическая повестка структурирована так, что онлайн-настроения вполне могут затрагивать онлайн-аудиторию за счет более активных пользователей, распространяющих информацию по сетевому принципу [18; 19; 20; 21]. С позиций методологической строгости можно рассматривать интернет-аудиторию не как плохую выборочную, а как самостоятельную генеральную совокупность, изменения в поведении которой сигнализируют о динамике общего отношения к ситуации [22; 23]. В пользу такого отношения говорят ремарки уже процитированных авторов, рассматривающих цифровые медиа «не только как инструмент, но как предпосылку социальных изменений» [17, р. 194].

Социальные сети в каком-то смысле олицетворяют все четыре перечисленные альтернативы социологическим опросам. Они ярко иллюстрируют тезис о том, что не все мнения одинаково важны, и ограничивают наше внимание не некоторым представительным срезом, а только тем отношением, которое было высказано, причем именно в той пропорции, в которой оно было высказано. Последнее позволяет создать условия, напоминающие глубинное интервью: убежденные пользователи, которым есть, что сказать, делают это, и их высказывания не ограничиваются (надо напомнить, что даже открытые вопросы в анкете имеют стандартную постановку и строго отведенное место для ответа); пользователи, не имеющие явного отношения к проблеме, высказываются реже или не высказываются вовсе. Все это доступно миллионам граждан и может быть агрегировано на разных уровнях, что наглядно продемонстрировала система управления негативом «Инцидент Менеджмент», активно внедряемая

на уровне российских регионов [24]. Вместе с тем по крайней мере часть вышеозначенных проблем, как то обилие спам-ботов или социальное профилирование, вполне может найти техническое решение [25; 26].

Хотя академическое сообщество скептически воспринимает попытки «подмены» социологии веб-скрейпингом, с начала 2010-х гг. было опубликовано сразу несколько эмпирических работ, демонстрирующих корреляцию между «полем» и «сетью». Так, исследователи из системы университетов Вермонта показали, что правильно отфильтрованная коллекция твитов может предсказывать рейтинги президента США на квартал вперед [27]. Их коллеги из Университета Индианы зафиксировали статистически значимую связь между упоминаниями кандидатов в члены Палаты представителей и их последующим результатом на выборах [28]. До этого API Twitter'а неоднократно использовался для более [29; 30] или менее [31; 32] удачных электоральных прогнозов. Для виртуальной социоскопии пилотировались и другие платформы (программные интерфейсы американских технологических гигантов и их конкурентов [33; 34]), а площадкой для нее становились в т. ч. развивающиеся страны — от Бразилии [35] до Китая [36].

Последний случай несколько «выбивается» из общего ряда работ, отличительная черта которых — тот факт, что они проводились в относительно свободной медиасреде, то есть в демократиях разной степени устойчивости. В авторитарных обществах и ответы на вопросы социологов, и публикации в сети несут дополнительную нагрузку; отдельные предпочтения могут иска жаться вплоть до своей полной противоположности [37]. Китайский «кейс» в этом смысле представляется экстремальным, когда социологии по широкому кругу вопросов просто нет. В российском контексте, совмещающем формальную информационную свободу с различными механизмами (само)цензуры, неоднократно предпринимались попытки «очистить» общественное

мнение от внешних факторов. Большинство из этих попыток так или иначе касались президентских рейтингов: официальные цифры поддержки Владимира Путина ставились под сомнение, но затем подтверждалась с помощью т. н. списочных экспериментов (*list experiments*) [38; 39]. Отчасти это объяснялось эффектом сплочения, впервые возникшем в 2014 г. в связи с присоединением Крыма и повторившемся в схожих обстоятельствах в 2022 г. [40]. Специфика же нынешнего периода, судя по всему, заключается в повышенной ригидности: россияне готовы поменять свое отношение к Президенту под воздействием групповых стимулов, но мобилизовать их для еще большей поддержки чрезвычайно трудно [41].

Таким образом, при должной обработке³, социологические данные предлагают наилучшее приближение к истинной⁴ картине политических настроений, и такое положение дел до сих пор не смогла поменять ни одна сетевая метрика. Стоит ли, однако, говорить о том, что в регулярных опросах эта обработка не проводится, и полученные цифры страдают от большого числа смещений. Мы считаем, что в России образца 2020-х гг. подобные смещения — эффекты социальной желательности, присоединения к (воспринимаемому) большинству, «спирали молчания» — действуют в одном и том же направлении и для сетей, и для опросов,

³ Это может быть не только списочно-экспериментальный дизайн, но и какие-то более простые техники, скажем, комбинация косвенных вопросов, призванных обойти сенситивные формулировки.

⁴ Здесь мы невольно занимаем позицию поллстерского реализма, полагая, что у граждан все-таки существует унимодальное распределение политических предпочтений, и трудности с их измерением связаны с недоработками опросчика (а не, например, отсутствием рефлексии опрашиваемых по большинству актуальных вопросов). Мы понимаем, что излагаемые в опросах или сетях мнения могут быть продуктом самых разных, необязательно глубоких, когнитивных процессов [42]. Но придерживаемся «нейтримной» гэллаповской позиции, поскольку именно ее придерживаются организаторы массовых исследований, альтернативу которым мы тестируем.

а значит (опять же, при соответствующей обработке) динамика их результатов будет совпадать. Мы предлагаем рассматривать именно временную изменчивость показателей, поскольку, как мы обсуждали выше, точечные оценки на репрезентативной выборке и фокусированной интернет-аудитории вряд ли когда-либо совпадут; доля «позитива» о политике в сети почти никогда не равна проценту его поддержки в опросе. Но вполне разумно ожидать, что серия позитивных информационных поводов о Президенте найдет отражение и в реакции респондентов, и в комментариях пользователей.

Мы рассчитываем извлечь выгоду из уникальной ситуации, сложившейся в РФ к 2021–2022 гг. Консолидированный персонализм публичной политики, на который «наложился» внешний шок начала СВО, позволяет нам проверить сразу несколько гипотез об измерении общественного мнения в условиях ограничений.

- H_1 : Уровень поддержки Президента в опросах отличается от усредненной тональности сообщений о нем в сети.
 - H_{11} : При социально-демографическом перевзвешивании сетевой «выборки», разница между этими показателей окажется меньше.
 - H_{12} : После 24 февраля 2022 г. разница между этими показателями будет больше.
- H_2 : Временные ряды поддержки Президента в опросах и усредненной тональности сообщений о нем в сети коинтегрированы.
 - H_{21} : При социально-демографическом перевзвешивании сетевой «выборки», близость рядов окажется более выраженной.
 - H_{22} : После 24 февраля 2022 г. близость рядов будет менее выраженной.

Методология и методы исследования

Для проверки этих гипотез, необходимо предварительно операционализировать индикаторы «сети» и «поля», которые, по нашему предположению, лучшим образом отразят каждый из измерительных подходов. Если с опросной составляющей дело обстоит относительно просто — достаточно выгрузить цифры с открытых порталов поллстеров⁵ [43; 44; 45⁶], сделать подходящий срез публикационной активности, сопряженной с деятельностью Владимира Путина, кажется нетривиальной задачей. Во-первых, вычислительные мощности не позволяют нам одновременно анализировать миллионы постов и комментариев, ежемесячно появляющихся в контексте работы Президента, поэтому какую-то их часть неизбежно придется «отсеять». Во-вторых, существенный процент онлайн-упоминаний не имеет прямого отношения к Президенту и представляет собой новостной или агитационный «шум». Наконец, все публикации в сети представлены в виде разноразмерных текстов, которые предстоит токенизировать, лемматизировать и «очистить» для оценки тональности, а ее, в свою очередь, агрегировать по временным промежуткам, совпадающим (для сопоставимости) с датами проведения опросов.

Как упоминалось во Введении, мы ограничиваем поиск периодом 24.08.2021 – 24.08.2022, то есть 12-ю месяцами с условным центром в дате начала СВО. Стартовой точкой отбора постов является составление словаря, в который мы включили фамилию Президента, словосочетание его имени и отчества и наи-

⁵ Всего были выгружены результаты 113 массовых опросов, в ходе которых респондентам задавался вопрос о доверии Владимиру Путину. В качестве социологического индикатора мы используем т. н. индекс доверия, т. е. разницу долей «доверяющих» и «не доверяющих» Президенту. Этот подход адаптируется и для социальных сетей.

⁶ Данный материал создан и распространен средством массовой информации, признанным выполняющим функции иностранного агента.

менование должности⁷. Чтобы исключить попадание в такую «выборку», например, сообщений о президенте Академии телевидения, постфактум были удалены 128 тысяч строк, содержащих названия стран или публичных организаций в пределах двух слов от леммы «президент». Кроме того, мы сфокусировались на публикациях отдельных пользователей, имеющих по меньшей мере 50 друзей и больше одного просмотра у рассматриваемого поста, чтобы избежать, с одной стороны, попадания текстов от крупных изданий, а с другой — текстов от малоактивных аккаунтов (вероятно, веб-ботов). Парсинг публикаций по словарю осуществлялся с помощью полуавтоматического сервиса «Медиалогия СМ». Фильтрация спама отчасти выполнялась внутри системы, отчасти — при последующей обработке (подозрительными считались публикации, состоящие на более чем 50 % из стоп-слов).

Дополнительные фильтры (в частности, отказ от рассмотрения мессенджеров⁸, нерусскоязычных постов и постов, сделанных в зарубежных странах) позволили нам сохранить почти 1,6 млн текстов (если текст был нанесен на фото или произнесся в аудиовизуальном формате, система считывала его автоматически), а также техническую информацию о каждой публикации и ее авторе. Это составляет около 10 % от общего числа упоминаний Президента в сети за тот же период. За четыре дня

⁷ Полностью, запрос можно представить как логическое выражение (Президент | Путин*) | (Владимир* / 1 Владимирovich*).

⁸ Это самоограничение продиктовано трудностями с социально-демографическим портретированием авторов. Интересно, что несмотря на отсутствие каких бы то ни было других требований (кроме полноты публикаций за рассматриваемый период), финальный массив сформировали ровно две отечественные («ВКонтакте», «Одноклассники») и две зарубежные (Instagram*, TikTok) сети. Гипотезы, предполагающие перевзвешивание, тестировались только на материале ВК и ОК.

* Компания-владелец Instagram, Meta Platforms Inc., признана экстремистской организацией. Ее деятельность на территории страны запрещена.

с 24.02.2022 по 27.02.2022 было зафиксировано аномальное количество сообщений и «вбросов», поэтому они исключены из дальнейших процедур агрегирования как влиятельные наблюдения.

Оценка тональности производилась на дезагрегированных данных, то есть для каждой публикации в отдельности. С помощью библиотеки семантического анализа Dostoevsky [46], предобученной на большом корпусе русскоязычных текстов, мы оцениваем вероятность принадлежности каждого поста к следующим категориям: «нейтральный», «негативный», «позитивный», «речевой акт», «другое». Заметим, что вероятность классификации поста как позитивного не является полным отражением позитивного восприятия пользователей, так как не учитывает распределения остаточных вероятностей: существует определенная разница между постом с 50-процентной вероятностью «позитива» и 5-процентной вероятностью «негатива» и постом с 50-процентной вероятностью «позитива» и 50-процентной вероятностью «негатива». Поэтому в качестве сетевой метрики настроений мы используем разницу

$$P_{Dostoevsky}[i\text{-я публикация} = \text{«позитивная»}] - \\ - P_{Dostoevsky}[i\text{-я публикация} = \text{«негативная»}].$$

Чтобы не ограничивать себя единственной мерой, мы повторяем все тесты на другой переменной тональности: нейросети «Медиалогии» по умолчанию присваивают каждой публикации маркер «нейтральная», «негативная» или «позитивная», который также можно квантифицировать (как «0», «-1» и «1», соответственно) и усреднить по выбранным промежуткам.

Итого, нам необходимо сравнить средние двух переменных, каждая из которых представляет собой разницу долей. Статистически, функцию проверки различий может выполнить простой тест типа ANOVA: переменные измерены за идентичные временные периоды и подвержены одному и тому же контексту, а значит, нужды в дополнительных «контролях» быть не должно. На деле

же, особенно в сети, появляется целый ряд третьих факторов: оценки могут отличаться по онлайн-площадкам, популярности авторов, подвыборкам пользователей (с учетом взвешивания и без него) и конкретным датам (в частности, до и после начала СВО). Еще одной опосредующей переменной является фактор поллстера: цифра одобрения может разниться в зависимости от формулировок вопроса и общего качества работы интервьюеров. Наконец, следует учесть сезонно-трендовую динамику, добавив в уравнение лаги наших замеров⁹. Приняв это во внимание, мы последовательно оцениваем серию регрессий:

$$\begin{cases} Support_Rate_{it} = \beta_0 + \beta_1 Audience_{it} + \beta_2 Involvement_{it} + \varepsilon_{it} & (1) \\ (Support_Rate_{it} - \widehat{Support_Rate}_{it}) = \gamma_0 + \gamma_1 Source_{it} + \gamma_2 Support_Rate_{it-1} \\ \quad [+ \gamma_3 Support_Rate_{it-2} + \gamma_4 SMO_{it} + \gamma_5 SMO_{it} Source_{it}] + e_{it} & (2), \end{cases}$$

где роль ключевого предиктора играет *Source*, отражающий принадлежность источника к поллстерам, оценкам по «Медиалогии» или оценкам по Dostoevsky, а ковариатами выступают средняя аудитория авторов в t -й период, средняя вовлеченность читателей (сумма «лайков», репостов и комментариев) в t -й период и лаги отклика. Сначала, с целью «очистить» целевой показатель от влияния онлайн-среды, оцениваются остатки модели на аудиторию и вовлеченность. Затем эти остатки регрессируются на *Source* либо его более подробную версию (*Source_Detailed*) с «разбивкой» по отдельным социологическим службам и социальным сетям. Чтобы проверить первую частную гипотезу, мы перевзвешиваем наблюдения (-публикации), добиваясь представительности по полу, возрасту и региону проживания авторов, и заново строим модель на *Source*. Для проверки второй частной гипотезы, мы добавляем в выражение дамми *SMO*, разделяющую

⁹ Мы выбрали этот способ как конвенциональный в нивелировке проблем, вызванных последовательно скоррелированными ошибками (serially correlated errors). Порядок лага определяется частной автокорреляционной функцией и может отличаться между моделями.

строки на «до» и «после» 24.02.2022, и переоцениваем уравнение с учетом ее взаимодействия с *Source*. В качестве последнего шага мы также добавляем взаимодействие в модель на взвешенных данных. При этом во всех спецификациях, независимо от присутствия дамми и взвешивания, рассчитываются устойчивые к гетероскедастичности стандартные ошибки Дрисколла-Края.

Тем самым, несмотря на все усилия по «контролю» за временем, мы моделируем разницу в статике, не будучи до конца уверены в динамическом соотношении переменных. Проверить гипотезы H_0 , мы рассчитываем с помощью коинтеграционного анализа, применения к рядам опросных и сетевых оценок МНК-тест Энгла-Грэнджа и авторегрессионный тест Йохансена. В основе первого лежит предположение, что взаимосвязь двух интегрируемых рядов выражается в стационарности остатков, полученных из линейной регрессии одного на другой. Тогда регрессию можно оценить по методу наименьших квадратов, а стационарность — протестировать с помощью статистики Дики-Фуллера. В рамках второго теста строится более сложная модель — векторная авторегрессия с коррекцией ошибок, матрица коэффициентов которой не должна оказаться нулевой. Проверить, так ли это, помогает статистика отношения правдоподобия (т. н. «статистика следа»). Мы реализуем обе процедуры для линейного тренда «сети» и «поля». Каждая из них тестируется против нулевой гипотезы об отсутствии связи и позволяет понять, можно ли составить из рядов стационарную линейную комбинацию, то есть, с учетом поправок, использовать как «заменители».

Результаты моделирования

Перед тем, как комментировать модели и коэффициенты, обратим внимание на целевую переменную *Support_Rate*. Ее опросная составляющая менялась сообразно «сплочению вокруг флага» и зарегистрировала закономерный рост доверия

после 24 февраля. «Скачок» индекса с ~30% до ~65% был оперативно отражен в инфографике всех социологических агентств. К сожалению, аналогичный график для сетевых метрик малоинформативен: из-за повышенной чувствительности к контексту, он напоминает серию локальных трендов с пилообразным блужданием вокруг них. Поэтому мы визуализируем динамику трех субкомпонент отклика (оценок полистеров, «Медиалогии» и Dostoevsky) loess-кривыми: на Рисунках 1 и 2 изображены сглаженные тенденции, значения которых получены с помощью взвешенных квадратичных регрессий.

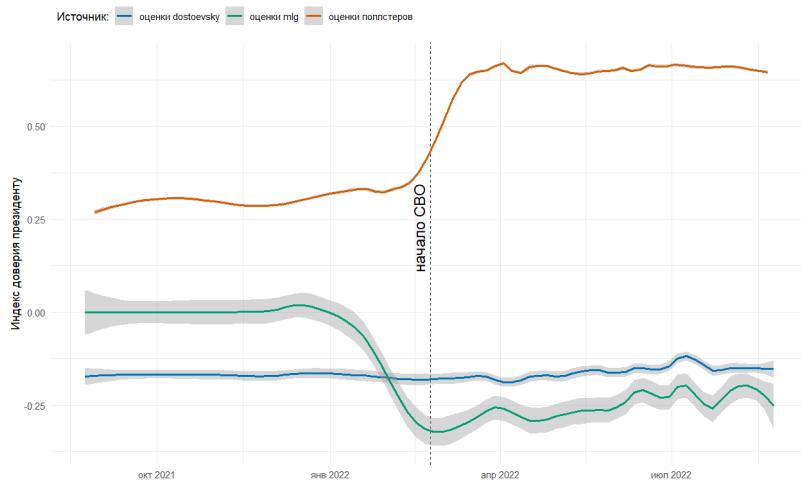


Рис. 1. Индекс доверия Владимиру Путину в социальных сетях и социологических опросах (исходные данные, динамика 24 августа 2021 – 24 августа 2022 гг.)

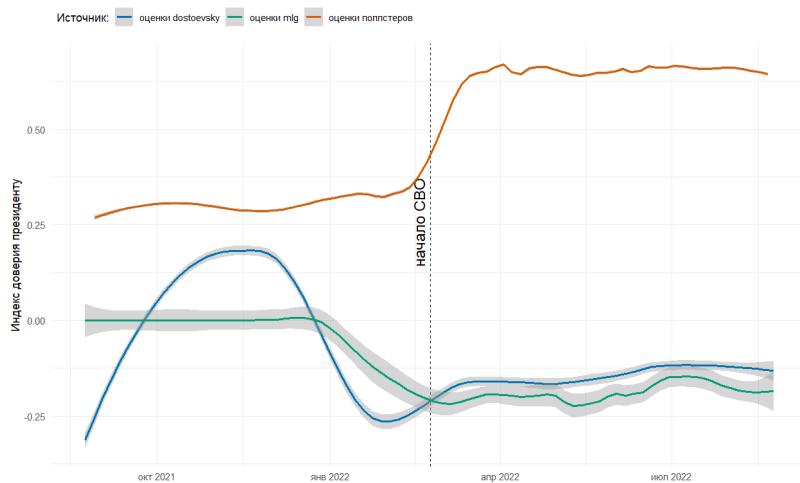


Рис. 2. Индекс доверия Владимиру Путину в социальных сетях и социологических опросах (взвешенные данные, динамика 24 августа 2021 – 24 августа 2022 гг.)

Невооруженным глазом видно, что пользователи сети выражают гораздо меньше доверия Президенту, чем участники опросов. Обстановка военной операции способствовала дальнейшему расщеплению «оффлайна»/«онлайна» и росту разброса в последнем. На этом фоне привлекает внимание малоподвижная функция Dostoevsky, большинство оценок которого сосредоточены в диапазоне -15 ± 3 . Невысокая дисперсия в них обусловлена, прежде всего, обилием наблюдений, доступных вплоть до минуты конкретного дня, а также большей детальностью по сравнению с «Медиалогией». Вдобавок квадратичное выравнивание, визуально нивелирующее шумы, скрывает ряд важных, но распределенных во времени факторов (которые могут усиливать разнонаправленность «сети» и «поля»). Так, если зафиксировать динамические изменения как разницу (вернее, прирост) между

периодами до и после начала СВО, окажется, что северные и южные регионы по-разному высказывались об этом в Рунете (Рис. 3). Отклонения от условной границы проявляются в самых неожиданных местах, например, на приграничных территориях и в национальных республиках. Такая дифференциация, не совпадающая с конвенциональными географическими расколами, в очередной раз подчеркивает специфику российского «онлайна».

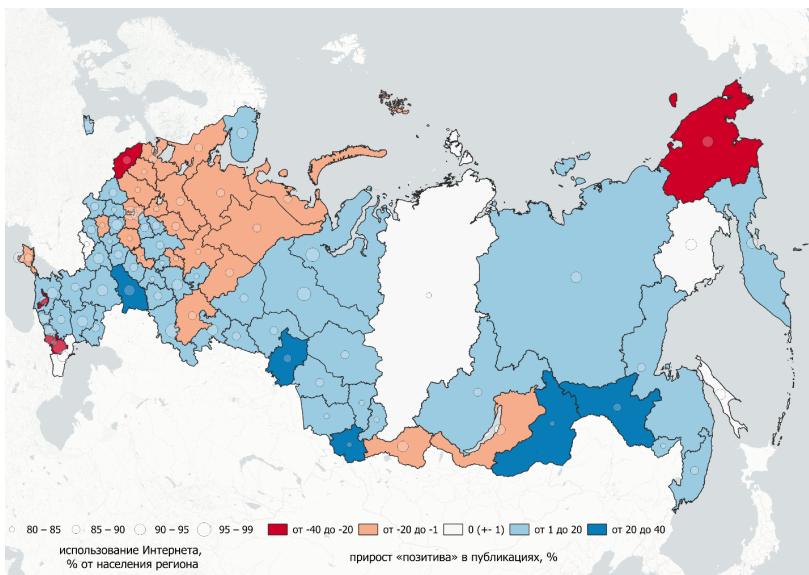


Рис. 3. Динамика тональности упоминаний Владимира Путина в социальных сетях (разница усредненных оценок Dostoevsky за полгода до и после начала СВО)

Оценки коэффициентов (Таблицы 1 и 2) подтверждают особенности онлайн-измерений. Относительно «базисных» данных опросов, социальные сети стабильно занижают индекс доверия Президенту на ~20 п. п. Взвешивание делает этот результат лишь

более выраженным. После начала боевых действий цифры расходятся: тональность по Dostoevsky почти не заметила изменений, а вот «Медиалогия» зафиксировала увеличение разрыва, который до СВО не был таким существенным. Стоит оговориться, что алгоритмы «Медиалогии» склонны маркировать сообщения как нейтральные, так что сравнение происходит с заниженной базой (это объясняет, почему до операции mlg-оценки находились ближе всего к опросным, а после резко сместились в отрицательную зону). Попытки выделить источники негатива (площадки, которые, предположительно, усугубляют разницу «сети» и «поля»), обрачиваются успехом только в период после 24 февраля. В Instagram и «Одноклассниках» конфликт встретили более критично, и второе полугодие отметилось для их пользователей снижением доверия к Президенту. В TikTok и «ВКонтакте» доверие также снизилось, но в меньшей степени. При этом исходные оценки у всех четырех платформ были ниже, чем у любой из социологических служб.

Таблица 1
ОЦЕНКИ ОСНОВНЫХ МОДЕЛЕЙ НА ИСХОДНОМ
И ВЗВЕШЕННОМ МАССИВАХ

Предиктор / Выборка	Отклик: Support_Rate			
	(исход.)	(взвеш.)	(исход.)	(взвеш.)
(Intercept)	0.183** (0.073)	0.204*** (0.078)	0.149*** (0.039)	0.148*** (0.040)
	-0.173** (0.072)	-0.196*** (0.070)	-0.180*** (0.037)	-0.203*** (0.031)
Source (pollster vs. dostoevsky)	-0.193*** (0.071)	-0.212*** (0.076)	-0.074*** (0.027)	-0.053* (0.029)
	0.640*** (0.153)	0.585*** (0.192)	0.533*** (0.127)	0.382*** (0.129)

Окончание табл. 1

Отклик: Support_Rate				
Предиктор / Выборка	(исход.)	(взвеш.)	(исход.)	(взвеш.)
Support_Rate _{t-2}		0.016		0.152***
		(0.071)		(0.045)
SMO			0.160***	0.162***
			(0.042)	(0.044)
Source (pollster vs. dostoevsky) × SMO			-0.082	-0.039
			(0.052)	(0.056)
Source (pollster vs. mlg) × SMO			-0.323***	-0.365***
			(0.083)	(0.077)
N	1010	555	1010	555
R ²	0.745	0.755	0.845	0.858
AIC	-1720.9	-739.7	-2216.4	-1035.9
BIC	-1696.3	-713.8	-2177.0	-997.0
SE	Driscoll & Kraay	Driscoll & Kraay	Driscoll & Kraay	Driscoll & Kraay

Примечание: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Таблица 2
ОЦЕНКИ МОДЕЛЕЙ НА «ПОДРОБНУЮ» ВЕРСИЮ
ОСНОВНОГО ПРЕДИКТОРА

		Отклик: Support_Rate	
Предиктор \ Выборка		(исход.)	(исход.)
(Intercept)		0.234***	0.283***
		(0.079)	(0.053)
Source_Detailed (levada vs. vciom)		-0.018***	0.018*
		(0.005)	(0.010)

Продолжение табл. 2

	Отклик: Support_Rate	
Предиктор \ Выборка	(исход.)	(исход.)
Source_Detailed (levada vs. fom)	-0.035*** (0.003)	-0.058*** (0.007)
	-0.248*** (0.084)	-0.230*** (0.055)
Source_Detailed (levada vs. tiktok)	-0.206*** (0.067)	-0.252*** (0.046)
	-0.220*** (0.077)	-0.240*** (0.056)
Source_Detailed (levada vs. vk)	-0.262*** (0.076)	-0.238*** (0.066)
	0.583*** (0.169)	0.146 (0.179)
SMO		0.269*** (0.055)
		-0.016*** (0.004)
Source_Detailed (levada vs. fom) × SMO		0.055*** (0.011)
		-0.424*** (0.099)
Source_Detailed (levada vs. instagram) × SMO		-0.221*** (0.045)
		-0.297*** (0.062)
Source_Detailed (levada vs. ok) × SMO		-0.463*** (0.108)

Окончание табл. 2

	Отклик: Support_Rate	
Предиктор \ Выборка	(исход.)	(исход.)
N	1010	1010
R ²	0.752	0.821
AIC	-1741.3	-2055.5
BIC	-1697.1	-1976.8
SE	Driscoll & Kraay	Driscoll & Kraay

Примечание: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Таким образом, как бы мы ни группировали источники, обнаружить совпадение их центральных тенденций не удается. Но, может быть, мы вправе говорить о совместной изменчивости? Результаты коинтеграционных тестов в Таблице 3 противоречивы. Процедура Энгла-Грэнджа не позволяет отвергнуть гипотезу о наличии единичного корня, т. е. об отсутствии связи. Тест Йохансена, напротив, намекает на возможность формирования стационарного ряда из попарных комбинаций с любым из сетевых измерений, кроме невзвешенных оценок Dostoevsky. Сделать однозначный вывод и предпочесть какой-либо из этих результатов нельзя. С одной стороны, двухшаговая методика Энгла-Грэнджа чрезмерно полагается на МНК и маломощный тест Дики-Фуллера; с другой, статистики Йохансена справедливы только асимптотически и создают положительное смещение (positive bias) на конечных выборках. В отсутствие других свидетельств, мы склоняемся именно к последней версии, хотя и не можем утверждать ее наверняка. Разбиение рядов по линии начала СВО дает схожую картину (Таблица 4): EG-тест не распознает никакой связи, а J-тест, в особенности примененный к взвешенным данным после 24.02.2022, оценивает порядок коинтеграции > 1 . Это можно осторожно трактовать в пользу гипотезы H_{21} , но против H_{22} .

Таблица 3

ОЦЕНКИ ТЕСТОВ ЭНГЛА-ГРЭНДЖЕРА И ЙОХАНСЕНА.
г ОТРАЖАЕТ ПРЕДПОЛАГАЕМЫЙ РАНГ КОИНТЕГРАЦИИ

	Ряд-отклик: pollster		
24.08.21 - 24.08.22	EG-тест	J-тест ($r = 0$)	J-тест ($r \leq 1$)
dostoevsky (исход.)	1.9	4.38	1.10
mlg (исход.)	1.12	21.79**	1.30
dostoevsky (взвеш.)	1.68	37.72***	0.79
mlg (взвеш.)	1.03	24.72***	0.87

Примечание: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Таблица 4

ОЦЕНКИ ТЕСТОВ ЭНГЛА-ГРЭНДЖЕРА И ЙОХАНСЕНА
НА ПОДМАССИВАХ ДО И ПОСЛЕ НАЧАЛА СВО.
г ОТРАЖАЕТ ПРЕДПОЛАГАЕМЫЙ РАНГ КОИНТЕГРАЦИИ

	Ряд-отклик: pollster		
24.08.21 - 24.02.22	EG-тест	J-тест ($r = 0$)	J-тест ($r \leq 1$)
dostoevsky (исход.)	1.59	18.28**	7.66*
mlg (исход.)	0.9	13.88	0.29
dostoevsky (взвеш.)	1.4	33.15***	7.74*
mlg (взвеш.)	0.96	13.74	0.34
24.02.22 - 24.08.22			
dostoevsky (исход.)	3.43	16.56*	1.21
mlg (исход.)	3.7	16.89*	0.97
dostoevsky (взвеш.)	2.92	35.97***	12.31***
mlg (взвеш.)	3.28	30.33***	13.29***

Примечание: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Заключение

В условиях дороговизны полевой социологии и неизбежных смещений, возникающих при ее применении к острополитическим вопросам, все большую важность приобретает поиск альтернативных инструментов изучения общественного мнения. Хотя вопрос о замене анкетных исследований, как правило, не стоит, интерес представляет сама возможность каким-то образом уточнить цифры опросов, не прибегая к проведению других опросов. Одним из самых распространенных дополнений к привычному «полю» является т. н. sentiment analysis — анализ настроений в информационном потоке социальных медиа. Если предположить, что интересующие нас тренды затрагивают и онлайн-, и офлайн-аудиторию (то есть неравенство доступа к глобальной сети не связано с предметом исследования), а онлайн-авторов можно локализовать (в частности, по принадлежности к стране), интернет превращается в своеобразную «большую выборку», которую с некоторыми корректировками можно сопоставить выборке опросной.

Достаточно ли (квази)случайного отбора авторов и их социально-демографического взвешивания для того, чтобы сделать интернет-выборку ближе к стандартной омнибусной? Наши модели, основанные на сравнении доверия Президенту в сети и в «поле», утверждают, что нет. Ни рассмотрение отдельных поллеров/площадок, ни изменение метода маркировки постов, ни коррекция выборки пользователей на веса из общероссийской переписи не позволяют сократить разрыв в 20 п. п., переходящий из модели в модель (именно на столько Рунет, предположительно, хуже воспринимает действия Владимира Путина). Начало украинской спецоперации как минимум не помогло сблизить оценки, а в некоторых случаях, наоборот, усугубило разницу между неизменно позитивными докладами социологов и крайне переменчивым сентиментом сети.

Можем ли мы, зафиксировав эту разницу, считать ее константой, а показатели в сети и в «поле» — соизменчивыми во времени? Ответом здесь, вероятнее всего, тоже будет «нет», хотя и менее уверенное. В серии тестов мы несколько раз фиксируем значимую коинтеграцию показателей, особенно явную при перевзвешивании сетевой выборки. Результаты всей серии, однако, противоречивы. Учитывая высокодисперсные настроения Рунета, в ряду которых могла возникнуть ложная автокорреляция, а также чувствительность используемого J-теста к ошибкам спецификации, мы можем назвать наблюдаемую значимость в лучшем случае «заделом», но никак не доказательством долговременной связи. В пользу такого вывода говорят и контринтуитивные результаты после февраля 2022 г.: вопреки ожиданиям, «опросное» доверие демонстрирует общие стохастические тренды со всеми измерениями «сетевого», в т. ч. с теми, отрыв от которых после начала СВО увеличился (с оценками «Медиалогии»).

Сходный содержательный вывод был сделан и в других работах, посвященных взаимной комплементарности opinion mining и opinion polling [47; 48]. Систематический перекос онлайн-мнений в пользу негатива [см. также 49; 50], отчасти вызванный невозможностью выразить накопившийся негатив офлайн, перекрывает ту выгоду, которую свобода коммуникаций в интернете имеет в сравнении с типовыми шкалами опросов. Специфика конкретных платформ, непрозрачность информации о пользователях и ограничения в обработке естественного языка делают перекрестную проверку «поля» «сетью» очень отдаленной перспективой. Ситуацию вряд ли изменит и «правильный» выбор сети. Как можно видеть, зарубежные развлекательные площадки (Instagram, TikTok) мало чем отличаются от российских универсальных («ВКонтакте», «Одноклассники»); негативные политические сюжеты чаще обсуждаются в заблокированных сетях со слабой модерацией, откуда происходит отток политически лояльных подписчиков, но этот фактор вполне можно проконтролировать.

Осторожный оптимизм внушают иные, не оговоренные до сих пор ограничения нашей работы. Во-первых, обнаруженные различия характерны только для президентского рейтинга, пожалуй, наиболее сенситивного показателя из возможных. В качестве менее «тефлоновых» индикаторов, можно было бы использовать доверие правительству или официальным лицам, достаточно узнаваемым в Рунете. Во-вторых, оценка различий (20 п. п. «отрыва») может меняться со временем. Хотя мы учтываем, что данные предстают в виде ряда, и не считаем, что политические симпатии россиян подвержены сезонным колебаниям, вероятно, выбор более широкого горизонта или многократное сэмплирование случайных периодов для анализа сделают разрыв менее выраженным. Наконец, делу приближения интернет-данных к опросным помогает развитие математической лингвистики. Уже опробованные «машины» демографического профилирования [51] и непараметрические классификаторы тональности [52] можно переложить на русскоязычную среду, и, нормируя дискуссию в блогах с целью отсечения выбросов (аномальной активности), получить менее смещенные оценки. В асимптотике, при увеличении числа пользователей до числа граждан, они будут все точнее отражать распространенность тех или иных взглядов, во всяком случае, тех, которые выражены публично.

ЛИТЕРАТУРА

1. Barnes J. E. Russian Public Appears to Be Souring on War Casualties, Analysis Shows // The New York Times [сайт]. 2023. URL: <https://www.nytimes.com/2023/05/26/us/politics/russia-public-opinion-ukraine-war.html> (дата обращения: 29.10.2025).
2. Kizilova K., Norris P. Assessing Russian Public Opinion on the Ukraine War // Russian Analytical Digest. 2022, № 281. P. 2-5. DOI: 10.3929/ethz-b-000539633.
3. Volkov D., Rosenfeld B., Morris J., Pleines H., Biriukova A., Koneva E., Chilingaryan A., Minailo A., Kamalov E., Sergeeva I., Zavadskaya M., Kostenko V. The Value of Public Opinion Polls // Russian Analytical Digest. 2023, № 292. P. 1-21. DOI: 10.3929/ethz-b-000599408.

4. *Best S. J., Krueger B. S.* Analyzing the Representativeness of Internet Political Participation // *Political Behavior*. 2005, vol. 27. P. 183-216. DOI: 10.1007/S11109-005-3242-Y.
5. *Chang L., Krosnick J. A.* National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality // *Public Opinion Quarterly*. 2009, vol. 73, № 4. P. 641-678. DOI: 10.1093/poq/nfp075.
6. *Smetanin S.* The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives // *IEEE Access*. 2020, vol. 8. P. 110693-110719. DOI: 10.1109/ACCESS.2020.3002215.
7. *Manza J., Brooks C.* How Sociology Lost Public Opinion: A Genealogy of a Missing Concept in the Study of the Political // *Sociological Theory*. 2012, vol. 30, № 2. P. 89-113. DOI: 10.1177/0735275112448054.
8. *Greene S. A., Robertson G. B.* *Putin v. The People: The Perilous Politics of a Divided Russia*. New Haven: Yale University Press, 2019. 287 p. DOI: 10.2307/j.ctvfc5417.
9. *Berinsky A. J.* Measuring Public Opinion with Surveys // *Annual Review of Political Science*. 2017, vol. 20, № 1. P. 309-329.
10. *Glynn C. J., Herbst S., Lindeman M., O'Keefe G. J., Shapiro R. Y.* Methods for Studying Public Opinion // *Public Opinion*. New York: Routledge, 2018. P. 57-86. DOI: 10.4324/9780429493256.
11. *Larsen E. G., Fazekas Z.* Alternatives to Opinion Polls: No Polls, Vox Pop, Poll Aggregators and Social Media // *Reporting Public Opinion: How the Media Turns Boring Polls into Biased News*. Cham: Palgrave Macmillan, 2021. P. 109-121. DOI: 10.1007/978-3-030-75350-4_6.
12. *Ravi K., Ravi V.* A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications // *Knowledge-Based Systems*. 2015, vol. 89. P. 14-46. DOI: 10.1016/j.knosys.2015.06.015.
13. *Ceron A., Curini L., Iacus S. M., Porro G.* Every Tweet Counts? How Sentiment Analysis of Social Media Can Improve Our Knowledge of Citizens' Political Preferences with an Application to Italy and France // *New Media & Society*. 2014, vol. 16, № 2. P. 340-358. DOI: 10.1177/1461444813480466.
14. *Stukal D., Sanovich S., Bonneau R., Tucker J. A.* Detecting Bots on Russian Political Twitter // *Big data*. 2017, vol. 5, № 4. P. 310-324. DOI: 10.1089/big.2017.0038.
15. Медиапотребление и активность в интернете // ВЦИОМ: [сайт]. 23.09.2021. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor-mediapotreblenie-i-aktivnost-v-internete> (дата обращения: 29.10.2025).
16. *Gunter B., Koteyko N., Atanasova D.* Sentiment Analysis: A Market-Relevant and Reliable Measure of Public Feeling? // *International Journal of Market Research*. 2014, vol. 56, № 2. P. 231-247. DOI: 10.2501/IJMR-2014-014.

17. Glazunova S. The “Sovereign Internet” and Social Media // Digital Activism in Russia: The Communication Tactics of Political Outsiders. Cham: Springer, 2022. P. 67-88. DOI: 10.1007/978-3-030-93503-0_4.
18. Skoric M. M., Zhu Q., Goh D., Pang N. Social Media and Citizen Engagement: A Meta-Analytic Review // New Media & Society. 2016, vol. 18, № 9. P. 1817-1839. DOI: 10.1177/1461444815616221.
19. Bisbee J., Larson J. M. Testing Social Science Network Theories with Online Network Data: An Evaluation of External Validity // American Political Science Review. 2017, vol. 111, № 3. P. 502-521. DOI: 10.1017/S0003055417000120.
20. Chae Y., Lee S., Kim Y. Meta-Analysis of the Relationship between Internet Use and Political Participation: Examining Main and Moderating Effects // Asian Journal of Communication. 2019, vol. 29, № 1. P. 35-54. DOI: 10.1080/01292986.2018.1499121.
21. Glenski M., Weninger T. Rating Effects on Social News Posts and Comments // ACM Transactions on Intelligent Systems and Technology. 2017, vol. 8, № 6. P. 1-19. DOI: 10.1145/2963104.
22. Feezell J. T., Conroy M., Guerrero M. Internet Use and Political Participation: Engaging Citizenship Norms through Online Activities // Journal of Information Technology & Politics. 2016, vol. 1, № 2. P. 95-107. DOI: 10.1080/19331681.2016.1166994.
23. Schober M. F., Pasek J., Guggenheim L., Lampe C., Conrad F. G. Social Media Analyses for Social Measurement // Public Opinion Quarterly. 2016, vol. 80, № 1. P. 180-211. DOI: 10.1093/POQ%2FNFV048.
24. Filatova O., Chugunov A., Bolgov R. Transformation of the Electronic Participation System in Russia in the Early 2020s: Centralization Trends // International Conference on Topical Issues of International Political Geography. Cham: Springer, 2021. P. 309-319. DOI: 10.1007/978-3-031-20620-7_27.
25. Pekar V., Najafi H., Binner J. M., Swanson R., Rickard C., Fry J. Voting Intentions on Social Media and Political Opinion Polls // Government Information Quarterly. 2022, vol. 39, № 4. P. 101658. DOI: 10.1016/j.giq.2021.101658.
26. Ouni S., Fkhi F., Omri M. N. A Survey of Machine Learning-Based Author Profiling from Texts Analysis in Social Networks // Multimedia Tools and Applications. 2023, vol. 82, № 24. P. 36653–36686. DOI: 10.1007/s11042-023-14711-8.
27. Cody E. M., Reagan A. J., Dodds P. S., Danforth C. M. Public Opinion Polling with Twitter. arXiv preprint. 1608.02024. 2016. DOI: 10.48550/arXiv.1608.02024.
28. DiGrazia J., McKelvey K., Bollen J., Rojas F. More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior // PloS One. 2013, vol. 8, № 11. P. e79449. DOI: 10.1371/journal.pone.0079449.

29. Tumasjan A., Sprenger T., Sandner P., Welpe I. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment // Proceedings of the international AAAI conference on web and social media. 2010, vol. 4, № 1. P. 178-185. DOI: 10.1609/icwsm.v4i1.14009.
30. Vaccari C., Valeriani A., Barberá P., Bonneau R., Jost J. T., Nagler J., Tucker J. Social Media and Political Communication. A Survey of Twitter Users during the 2013 Italian General Election // Italian Political Science Review. 2013, vol. 43, № 3. P. 381-410. DOI: 10.1426/75245.
31. Gayo-Avello D. "I Wanted to Predict Elections with Twitter and All I Got Was This Lousy Paper" A Balanced Survey on Election Prediction Using Twitter Data. arXiv preprint. 1204.6441. 2012. DOI: 10.48550/arXiv.1204.6441.
32. Gayo-Avello D. A Meta-Analysis of State-of-the-Art Electoral Prediction from Twitter Data // Social Science Computer Review. 2013, vol. 31, № 6. P. 649-679. DOI: 10.1177/0894439313493979.
33. Drus Z., Khalid H. Sentiment Analysis in Social Media and Its Application: Systematic Literature Review // Procedia Computer Science. 2019, vol. 161. P. 707-714. DOI: 10.1016/j.procs.2019.11.174.
34. Chauhan P., Sharma N., Sikka G. The Emergence of Social Media Data and Sentiment Analysis in Election Prediction // Journal of Ambient Intelligence and Humanized Computing. 2021, vol. 12. P. 2601-2627. DOI: 10.1007/s12652-020-02423-y.
35. Oliveira D. J. S., Bermejo P. H. S., dos Santos P. A. Can Social Media Reveal the Preferences of Voters? A Comparison between Sentiment Analysis and Traditional Opinion Polls // Journal of Information Technology & Politics. 2016, vol. 14, № 1. P. 34-45. DOI: 10.1080/19331681.2016.1214094.
36. Cheng J., Zhang X., Li P., Zhang S., Ding Z., Wang H. Exploring Sentiment Parsing of Microblogging Texts for Opinion Polling on Chinese Public Figures // Applied Intelligence. 2016, vol. 45. P. 429-442. DOI: 10.1007/s10489-016-0768-0.
37. Maleki A. How Do Leading Methods Mislead? Measuring Public Opinions in Authoritarian Contexts // Proceedings of the IPSA. 2021. P. 1-31.
38. Frye T., Gehlbach S., Marquardt K. L., Reuter O. J. Is Putin's Popularity Real? // Post-Soviet Affairs. 2017, vol. 33, № 1. P. 1-15. DOI: 10.1080/1060586X.2016.1144334.
39. Frye T., Gehlbach S., Marquardt K. L., Reuter O. J. Is Putin's Popularity (Still) Real? A Cautionary Note on Using List Experiments to Measure Popularity in Authoritarian Regimes // Post-Soviet Affairs. 2023, vol. 39, № 3. P. 213-222. DOI: 10.1080/1060586X.2023.2187195.
40. Hale H. E. Authoritarian Rallying as Reputational Cascade? Evidence from Putin's Popularity Surge after Crimea // American Political Science Review. 2022, vol. 116, № 2. P. 580-594. DOI: 10.1017/S0003055421001052.

41. Buckley N., Marquardt K. L., Reuter O. J., Tertychnaya K. Endogenous Popularity: How Perceptions of Support Affect the Popularity of Authoritarian Regimes // American Political Science Review. 2024, vol. 118, № 2. P. 1046-1052. DOI: 10.1017/S0003055423000618.
42. Шуберт И. Действительно ли опросы общественного мнения изучают общественное мнение? // Социологические исследования. 2018, № 12. С. 56-62. DOI: 10.31857/S013216250003169-8.
43. Доверие политикам // ВЦИОМ: [сайт]. 2022. URL: <https://wciom.ru/ratings/doverie-politikam/> (дата обращения: 29.10.2025).
44. Доминанты. Поле мнений. Выпуск 33 // ФОМ: [сайт]. 2022. URL: <https://fom.ru/Dominanty/14770> (дата обращения: 29.10.2025).
45. Одобрение органов власти // Левада-центр: [сайт]. 2022. URL: <https://www.levada.ru/indikatory/odobrenie-organov-vlasti/> (дата обращения: 29.10.2025).
46. Bureaucratic Labs. Dostoevsky Sentiment analysis library for Russian language // Github: [сайт]. 2021. URL: <https://github.com/bureaucratic-labs/dostoevsky> (дата обращения: 29.10.2025).
47. Дудина В. И., Юдина Д. И. Извлекая мнения из сети Интернет: могут ли методы анализа текстов заменить опросы общественного мнения? // Мониторинг общественного мнения: экономические и социальные перемены. 2017, № 5. С. 63-78. DOI: 10.14515/monitoring.2017.5.05.
48. Оценка качества опроса в сложных реалиях сегодняшнего дня. Социальные медиа в исследовании общественного мнения: отчет рабочей группы AAPOR о новых технологиях в исследовании общественного мнения // Американская ассоциация исследователей общественного мнения / Пер. с англ. В. Л. Силаевой, О. А. Оберемко. М.: ВЦИОМ, 2017. 78 с.
49. Анкудинов И. А. Патриотический дискурс в Рунете: до и после 24 февраля 2022 г. // Мониторинг общественного мнения: экономические и социальные перемены. 2024, № 2. С. 99-123. DOI: 10.14515/monitoring.2024.2.2515.
50. Анташева М. С., Лобанова П. А., Исаева Ю. К., Сабидаева Е. А., Пиекалникис А. С., Логинова И. В. Сентимент-анализ как метод исследования информационной повестки и общественного мнения (на примере СМИ и социальных сетей КНР) // Социология: методология, методы, математическое моделирование (Социология: 4M). 2024, № 57. С. 7-41. DOI: 10.19181/4m.2023.32.2.1.
51. Barberá P. Less Is More? How Demographic Sample Weights Can Improve Public Opinion Estimates Based on Twitter Data // NYU Working Paper. 2016. P. 1-37.
52. Hopkins D. J., King G. A Method of Automated Nonparametric Content Analysis for Social Science // American Journal of Political Science. 2010, vol. 54, № 1. P. 229-247. DOI: 10.1111/j.1540-5907.2009.00428.X.

Сведения об авторе

Анкудинов Иван Андреевич

Старший преподаватель, аспирант Национального
исследовательского университета «Высшая школа
экономики»; методолог Института Человека РОМИР
SPIN-код: 9709-0850
AuthorID: 1207349

DOI: 10.19181/4m.2025.34.2.4.

PROSPECTS OF MEDIA MONITORING IN PUBLIC OPINION RESEARCH (USING THE EXAMPLE OF TRUST IN THE PRESIDENT)

Ankudinov Ivan A.
HSE University, Moscow, Russia
iankudinov@hse.ru
ORCID: 0000-0002-7001-4163

For citation: Ankudinov I.A. Prospects of Media Monitoring in Public Opinion Research (Using the Example of Trust in the President). *Sotsiologiya: 4M (Sociology: methodology, methods, mathematical modeling)*, 2025, no. 61, p. 165-203.
DOI: 10.19181/4m.2025.34.2.4.

Abstract. The changing political mood of Russians is a constant subject of interest for sociological agencies. With the development of the Internet, conventional questionnaire research began to be supplemented by online surveys and, despite some skepticism, by social media mining. This article attempts to adjust an accidental web-sample so as to bring its estimates closer to representative omnibuses. We use media- and polling measures of trust in the president Putin to answer the questions: How far apart are these sources? Is it possible to bring their estimates closer using post-stratification weights? Has the relationship between them been disrupted by the onset of the Ukrainian conflict? Is this relationship – despite some absolute difference – stable over time, i.e. Is it possible to move from one indicator to another using linear transformations? The conclusions we reach based on the results of the statistical analysis are as follows: (1) Internet measurements based on average dictionary sentiment estimate trust in the president by 20 percentage points lower than surveys. (2) Adjusting the user sample for weights from the population census does not allow the estimates to be reconciled. (3) Clarification of estimates for separate networks and pollsters only confirms the detected difference. For some platforms this difference is more significant, which can be explained by the specifics of their audiences. (4) The “Ukrainian factor” had an ambiguous impact on the

discrepancy between estimates. Tonality according to “Medialogia” records an increase in the gap, while Dostoevsky sees almost no difference between “before” and “after”. (5) The temporal dependence of offline and online estimates is no less ambiguous. Weighting the Internet sample gives some reason to consider the corresponding time series to be cointegrated, but this relationship does not appear in all tests.

Keywords: pollsters, social media, weighting, opinion mining, text analysis, time series, presidential rating, public trust

References

1. Barnes J.E. “Russian Public Appears to Be Souring on War Casualties, Analysis Shows”, in: *The New York Times* [site]. 2023. URL: <https://www.nytimes.com/2023/05/26/us/politics/russia-public-opinion-ukraine-war.html> (date of access: 29.10.2025).
2. Kizilova K., Norris P. Assessing Russian Public Opinion on the Ukraine War, *Russian Analytical Digest*, 2022, no. 281, p. 2-5. DOI: 10.3929/ethz-b-000539633.
3. Volkov D., Rosenfeld B., Morris J., Pleines H., Biriukova A., Koneva E., Chilingaryan A., Minailo A., Kamalov E., Sergeeva I., Zavadskaya M., Kostenko V. The Value of Public Opinion Polls, *Russian Analytical Digest*, 2023, no. 292, p. 1-21. DOI: 10.3929/ethz-b-000599408.
4. Best S.J., Krueger B.S. Analyzing the Representativeness of Internet Political Participation, *Political Behavior*, 2005, vol. 27, p. 183-216. DOI: 10.1007/S11109-005-3242-Y.
5. Chang L., Krosnick J.A. National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality, *Public Opinion Quarterly*, 2009, vol. 73, no. 4, p. 641-678. DOI: 10.1093/poq/nfp075.
6. Smetanin S. The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives, *IEEE Access*, 2020, vol. 8, p. 110693-110719. DOI: 10.1109/ACCESS.2020.3002215.
7. Manza J., Brooks C. How Sociology Lost Public Opinion: A Genealogy of a Missing Concept in the Study of the Political, *Sociological Theory*, 2012, vol. 30, no. 2, p. 89-113. DOI: 10.1177/0735275112448054.

8. Greene S.A., Robertson G.B. *Putin v. The People: The Perilous Politics of a Divided Russia*. New Haven: Yale University Press, 2019. 287 p. DOI: 10.2307/j.ctvfc5417.
9. Berinsky A.J. Measuring Public Opinion with Surveys, *Annual Review of Political Science*, 2017, vol. 20, no. 1, p. 309-329.
10. Glynn C.J., Herbst S., Lindeman M., O'Keefe G.J., Shapiro R.Y. "Methods for Studying Public Opinion", in: *Public Opinion*. New York: Routledge, 2018, p. 57-86. DOI: 10.4324/9780429493256.
11. Larsen E.G., Fazekas Z. "Alternatives to Opinion Polls: No Polls, Vox Pop, Poll Aggregators and Social Media", in: *Reporting Public Opinion: How the Media Turns Boring Polls into Biased News*. Cham: Palgrave Macmillan, 2021, p. 109-121. DOI: 10.1007/978-3-030-75350-4_6.
12. Ravi K., Ravi V. A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications, *Knowledge-Based Systems*, 2015, vol. 89 p. 14-46. DOI: 10.1016/j.knosys.2015.06.015.
13. Ceron A., Curini L., Iacus S.M., Porro G. Every Tweet Counts? How Sentiment Analysis of Social Media Can Improve Our Knowledge of Citizens' Political Preferences with an Application to Italy and France, *New Media & Society*, 2014, vol. 16, no. 2, p. 340-358. DOI: 10.1177/1461444813480466.
14. Stukal D., Sanovich S., Bonneau R., Tucker J.A. Detecting Bots on Russian Political Twitter, *Big data*, 2017, vol. 5, no. 4, p. 310-324. DOI: 10.1089/big.2017.0038.
15. Media consumption and online activity (in Russian), in: *VCIOM* [site]. 23.09.2021. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor-mediapotreblenie-i-aktivnost-v-internete> (date of access: 29.10.2025).
16. Gunter B., Koteyko N., Atanasova D. Sentiment Analysis: A Market-Relevant and Reliable Measure of Public Feeling? *International Journal of Market Research*, 2014, vol. 56, no. 2, p. 231-247. DOI: 10.2501/IJMR-2014-014.
17. Glazunova S. "The "Sovereign Internet" and Social Media", in: *Digital Activism in Russia: The Communication Tactics of Political Outsiders*. Cham: Springer, 2022, p. 67-88. DOI: 10.1007/978-3-030-93503-0_4.
18. Skoric M.M., Zhu Q., Goh D., Pang N. Social Media and Citizen Engagement: A Meta-Analytic Review, *New Media & Society*, 2016, vol. 18, no. 9, p. 1817-1839. DOI: 10.1177/1461444815616221.

19. Bisbee J., Larson J.M. Testing Social Science Network Theories with Online Network Data: An Evaluation of External Validity, *American Political Science Review*, 2017, vol. 111, no. 3, p. 502-521. DOI: 10.1017/S0003055417000120.
20. Chae Y., Lee S., Kim Y. Meta-Analysis of the Relationship between Internet Use and Political Participation: Examining Main and Moderating Effects, *Asian Journal of Communication*, 2019, vol. 29, no. 1, p. 35-54. DOI: 10.1080/01292986.2018.1499121.
21. Glenski M., Weninger T. Rating Effects on Social News Posts and Comments, *ACM Transactions on Intelligent Systems and Technology*, 2017, vol. 8, no. 6, p. 1-19. DOI: 10.1145/2963104.
22. Feezell J.T., Conroy M., Guerrero M. Internet Use and Political Participation: Engaging Citizenship Norms through Online Activities, *Journal of Information Technology & Politics*, 2016, vol. 1, no. 2, p. 95-107. DOI: 10.1080/19331681.2016.1166994.
23. Schober M.F., Pasek J., Guggenheim L., Lampe C., Conrad F.G. Social Media Analyses for Social Measurement, *Public Opinion Quarterly*, 2016, vol. 80, no. 1, p. 180-211. DOI: 10.1093/POQ%2FNFV048.
24. Filatova O., Chugunov A., Bolgov R. “Transformation of the Electronic Participation System in Russia in the Early 2020s: Centralization Trends”, in: *International Conference on Topical Issues of International Political Geography*. Cham: Springer, 2021, p. 309-319. DOI:10.1007/978-3-031-20620-7_27.
25. Pekar V., Najafi H., Binner J. M., Swanson R., Rickard C., Fry J. Voting Intentions on Social Media and Political Opinion Polls, *Government Information Quarterly*, 2022, vol. 39, no. 4, p. 101658. DOI: 10.1016/j.giq.2021.101658.
26. Ouni S., Fkih F., Omri M.N. A Survey of Machine Learning-Based Author Profiling from Texts Analysis in Social Networks, *Multimedia Tools and Applications*, 2023, vol. 82, no. 24, p. 36653–36686. DOI: 10.1007/s11042-023-14711-8.
27. Cody E.M., Reagan A.J., Dodds P.S., Danforth C.M. *Public Opinion Polling with Twitter*. arXiv preprint. 1608.02024. 2016. DOI: 10.48550/arXiv.1608.02024.
28. DiGrazia J., McKelvey K., Bollen J., Rojas F. More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior,

- PloS One*, 2013, vol. 8, no. 11, p. e79449. DOI: 10.1371/journal.pone.0079449.
29. Tumasjan A., Sprenger T., Sandner P., Welpe I. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment, *Proceedings of the international AAAI conference on web and social media*, 2010, vol. 4, no. 1, p. 178-185. DOI: 10.1609/icwsm.v4i1.14009.
30. Vaccari C., Valeriani A., Barberá P., Bonneau R., Jost J.T., Nagler J., Tucker J. Social Media and Political Communication. A Survey of Twitter Users during the 2013 Italian General Election, *Italian Political Science Review*, 2013, vol. 43, no. 3, p. 381-410. DOI: 10.1426/75245.
31. Gayo-Avello D. “*I Wanted to Predict Elections with Twitter and All I Got Was This Lousy Paper*” A Balanced Survey on Election Prediction Using Twitter Data. arXiv preprint. 1204.6441. 2012. DOI: 10.48550/arXiv.1204.6441.
32. Gayo-Avello D. A Meta-Analysis of State-of-the-Art Electoral Prediction from Twitter Data, *Social Science Computer Review*, 2013, vol. 31, no. 6, p. 649-679. DOI: 10.1177/0894439313493979.
33. Drus Z., Khalid H. Sentiment Analysis in Social Media and Its Application: Systematic Literature Review, *Procedia Computer Science*, 2019, vol. 161, p. 707-714. DOI: 10.1016/j.procs.2019.11.174.
34. Chauhan P., Sharma N., Sikka G. The Emergence of Social Media Data and Sentiment Analysis in Election Prediction, *Journal of Ambient Intelligence and Humanized Computing*, 2021, vol. 12, p. 2601-2627. DOI: 10.1007/s12652-020-02423-y.
35. Oliveira D.J.S., Bermejo P.H.S., dos Santos P.A. Can Social Media Reveal the Preferences of Voters? A Comparison between Sentiment Analysis and Traditional Opinion Polls, *Journal of Information Technology & Politics*, 2016, vol. 14, no. 1, p. 34-45. DOI: 10.1080/19331681.2016.1214094.
36. Cheng J., Zhang X., Li P., Zhang S., Ding Z., Wang H. Exploring Sentiment Parsing of Microblogging Texts for Opinion Polling on Chinese Public Figures, *Applied Intelligence*, 2016, vol. 45, p. 429-442. DOI: 10.1007/s10489-016-0768-0.
37. Maleki A. How Do Leading Methods Mislead? Measuring Public Opinions in Authoritarian Contexts, *Proceedings of the IPSA*, 2021, p. 1-31.
38. Frye T., Gehlbach S., Marquardt K.L., Reuter O.J. Is Putin’s Popularity

- Real? *Post-Soviet Affairs*, 2017, vol. 33, no. 1, p. 1-15. DOI: 10.1080/1060586X.2016.1144334.
39. Frye T., Gehlbach S., Marquardt K.L., Reuter O.J. Is Putin's Popularity (Still) Real? A Cautionary Note on Using List Experiments to Measure Popularity in Authoritarian Regimes, *Post-Soviet Affairs*, 2023, vol. 39, no. 3, p. 213-222. DOI: 10.1080/1060586X.2023.2187195.
40. Hale H.E. Authoritarian Rallying as Reputational Cascade? Evidence from Putin's Popularity Surge after Crimea, *American Political Science Review*, 2022, vol. 116, no. 2, p. 580-594. DOI: 10.1017/S0003055421001052.
41. Buckley N., Marquardt K.L., Reuter O.J., Tertychnaya K. Endogenous Popularity: How Perceptions of Support Affect the Popularity of Authoritarian Regimes, *American Political Science Review*, 2024, vol. 118, no. 2, p. 1046-1052. DOI: 10.1017/S0003055423000618.
42. Šubrt J. Do Public Opinion Polls actually Provide a Survey of Public Opinion? (in Russian), *Sociological Studies*, 2018, no. 12, p. 56-65. DOI: 10.31857/S013216250003169-8.
43. Trust in politicians (in Russian), in: *VCIOM* [site]. 2022. URL: <https://wciom.ru/ratings/doverie-politikam/> (date of access: 29.10.2025).
44. Dominants. Field of Opinions. Issue 33 (in Russian), in: *FOM* [site]. 2022. URL: <https://fom.ru/Dominanty/14770> (date of access: 29.10.2025).
45. Approval of authorities (in Russian), in: *Levada-center* [site]. 2022. URL: <https://www.levada.ru/indikatory/odobrenie-organov-vlasti/> (date of access: 29.10.2025).
46. Bureaucratic Labs. Dostoevsky Sentiment analysis library for Russian language, in: *Github* [site]. 2021. URL: <https://github.com/bureaucratic-labs/dostoevsky> (date of access: 29.10.2025).
47. Dudina V.I., Iudina D.I. Mining Opinions on the Internet: Can the Text Analysis Methods Replace Public Opinion Polls? (in Russian), *Monitoring of Public Opinion: Economic and Social Changes*, 2017, no. 5, p. 63-78. DOI: 10.14515/monitoring.2017.5.05.
48. "Evaluating Survey Quality in Today's Complex Environment. Social Media in Public Opinion Research: Executive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research" (transl., in Russian), in: *American Association for Public Opinion Research*. Moscow: VCIOM. 78 p.

49. Ankudinov I.A. Patriotic Discourse in Runet: Before and after February 24, 2022 (in Russian), *Monitoring of Public Opinion: Economic and Social Changes*, 2024, no. 2, p. 153-177. DOI: 10.14515/monitoring.2024.2.2515.
50. Antasheva M.S., Lobanova P.A., Isaeva I.K., Sabidaeva E.A., Piekalnits A.S., Loginova I.V. Sentiment analysis as an information agenda and public opinion research method (on the example of Chinese mass media and social networks) (in Russian), *Sociology: Methodology, Methods, Mathematical Modeling (Sociology: 4M)*, 2024, no. 2, p. 7-41. DOI: 10.19181/4m.2023.32.2.1.
51. Barberá P. Less Is More? How Demographic Sample Weights Can Improve Public Opinion Estimates Based on Twitter Data, *NYU Working Paper*, 2016, p. 1-37.
52. Hopkins D.J., King G. A Method of Automated Nonparametric Content Analysis for Social Science, *American Journal of Political Science*, 2010, vol. 54, no. 1, p. 229-247. DOI: 10.1111/J.1540-5907.2009.00428.X.

Information about the author

Ivan A. Ankudinov

Senior Lecturer, PhD Student at HSE University; Methodologist
at the ROMIR
SPIN-code: 9709-0850
AuthorID: 1207349