

---

## Цифровые технологии и большие данные в социологических исследованиях: концепция, методология, возможности

---

Д.Л. КОНСТАНТИНОВСКИЙ\*, Е.С. ПОПОВА\*\*, И.С. КУЗНЕЦОВ\*\*\*,  
Р.С. КУЗНЕЦОВ\*\*\*\*

\***Давид Львович Константиновский** – доктор социологических наук, руководитель Отдела социологии образования, Институт социологии ФНИСЦ РАН, Москва, Россия, scan21@mail.ru, <https://orcid.org/0000-0003-3316-0644>

\*\***Екатерина Сергеевна Попова** – кандидат социологических наук, ведущий научный сотрудник, Отдел социологии образования, Институт социологии ФНИСЦ РАН, Москва, Россия, [esporova@isras.ru](mailto:esporova@isras.ru), <https://orcid.org/0000-0002-9808-3152>

\*\*\***Игорь Сергеевич Кузнецов** – кандидат социологических наук, научный сотрудник, Отдел социологии образования, Институт социологии ФНИСЦ РАН, Москва, Россия, [kuznetsov.igor@gmail.com](mailto:kuznetsov.igor@gmail.com), <https://orcid.org/0000-0002-4477-5233>

\*\*\*\***Роман Сергеевич Кузнецов** – младший научный сотрудник, Отдел социологии образования, Институт социологии ФНИСЦ РАН, Москва, Россия, [roman@c-discurs.ru](mailto:roman@c-discurs.ru), <https://orcid.org/0000-0003-3063-9465>

**Цитирование:** Константиновский Д.Л., Попова Е.С., Кузнецов И.С., Кузнецов Р.С. (2025) Цифровые технологии и большие данные в социологических исследованиях: концепция, методология, возможности // Мир России. Т. 34. № 1. С. 144–160. DOI: 10.17323/1811-038X-2025-34-1-144-160

### Аннотация

*Процессы цифровизации преобразили многие сферы социальной жизни. По одним направлениям внедрение цифровых технологий проходило в плавном, «сглаженном» режиме, в других сферах пришлось экстренно осваивать новые цифровые технологии. Сфера социологических исследований не является исключением: цифровые технологии сбора и анализа данных прочно укореняются в практике исследований социальных наук. Их применение и анализ больших данных в социологических исследованиях выступают колоссальным ресурсом прорывных исследований социальной действительности, но и не меньшим риском в ее объяснении и разработке практических рекомендаций по решению социально-прикладных задач. Прежде чем приступить к работе с цифровыми технологиями, инструментами и большими данными, необходимо прояснить их теоретико-методологическую природу и потенциал практической значимости с позиции академической социальной науки.*

---

Авторский коллектив выражает благодарность П.М. Козыревой за поддержку в развитии направления, Е.И. Григорьевой за помощь и ценные советы, М.С. Косолапову за глубокие вопросы по концепции и методологии.

Статья поступила в редакцию в марте 2024 г.

Целью статьи является анализ возможностей, ограничений и первых уроков существующей к настоящему времени практики применения цифровых технологий (в том числе больших данных) в социологических исследованиях. Рассматриваются особенности сбора и анализа таких данных, анализируются различные способы их интерпретации. Внимание акцентируется на сопряжении социальных наук, больших данных и машинного обучения. В частности, отмечается, что до сих пор уделяется мало внимания моделям машинного обучения. Подчеркивается необходимость развивать математическую и алгоритмическую стороны указанного синтеза. Делается предположение, что умеренный эмпиризм и априоризм способны стать отправными точками для плодотворного обогащения методологии, эпистемологии и принципов получения социологических знаний на основе больших данных. Формулируются ответственность и роль социолога в решении социально-прикладных задач при анализе больших данных. Обосновывается критичность применения метода триангуляции данных.

**Ключевые слова:** социологические исследования, социологические методы, методология, анализ данных, цифровизация, цифровые технологии, методы машинного обучения, большие данные

«Будущее уже здесь.  
Просто оно еще неравномерно распределено»  
Уильям Гибсон

## Введение

С 2004 г. по настоящее время при обращении к аналитике Google Trends для изучения динамики популярности поискового запроса «большие данные» (*big data*) по всему миру в категории «Наука» мы видим экспоненциальный рост интереса с 2010 по 2016 г., интерес остается стабильно высоким и по сей день (*рисунок 1*). С этим поисковым запросом связаны такие темы, как цифровые технологии, машинное обучение, аналитика и анализ данных, искусственный интеллект, облачные вычисления и алгоритмы. Россия находится на 67 месте из 71.

Развитие цифровых технологий делает большие данные повсеместными<sup>1</sup>. Активность индивидов регистрируется как в онлайн-пространстве (покупки, лайки, посты, поисковые запросы), так и в обыденной жизни (видео/аудиорегистрации; корпоративные и правительственные базы данных; экономические, банковские и прочие транзакции; географические метки). Все эти данные собираются

---

<sup>1</sup> Понятие «большие данные» в свою очередь также не остается статичным. Еще совсем недавно большие данные характеризовались так называемым правилом «3V» (*volume, variety, velocity* – объем, разнообразие, скорость) – тремя определяющими свойствами или измерениями больших данных [Гурвич и др. 2015]. Впоследствии определение расширилось до пяти составляющих путем добавления изменчивости (*variability*) и ценности (*value*) данных. Под изменчивостью понимается различие в значениях одних и тех же данных в зависимости от контекста. Под ценностью подразумевается извлечение максимальной пользы из их анализа. В современной парадигме установилось правило «7V». Достоверность (*veracity*) – одна из важнейших характеристик данных. И, наконец, визуализация (*visualization*) данных, особенно на этапе знакомства с данными, понимания их «портрета», визуальной аналитики потоков данных, как неотъемлемая и необходимая часть анализа и понимания структуры больших данных.

и накапливаются, превращаясь в бесчисленные потоки непрерывной информации. Такое стремительное накопление больших данных и кажущаяся легкость их сбора стимулируют появление различных подходов к использованию полученной информации в интерпретации социальных явлений и процессов.

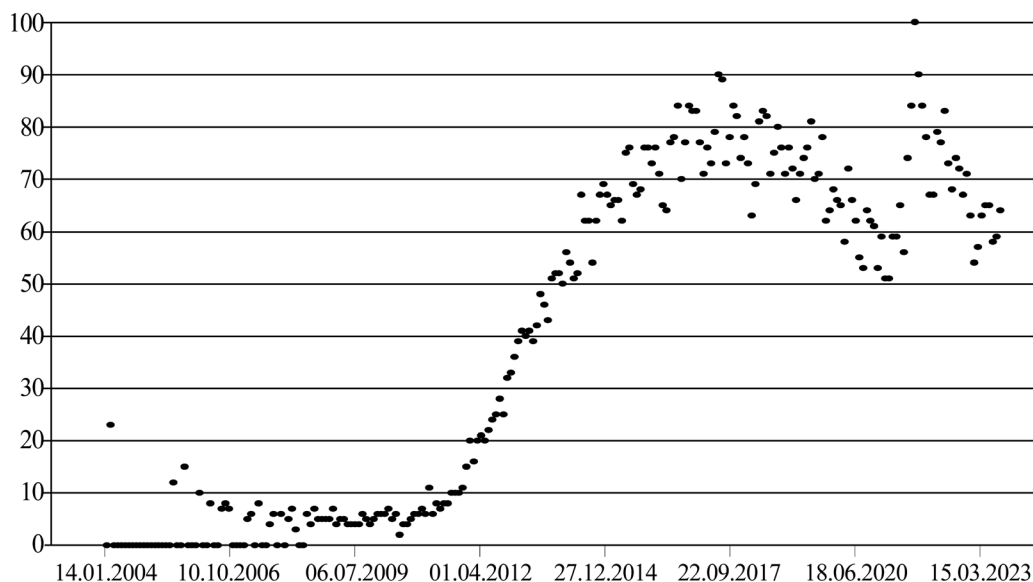


Рисунок 1. Динамика интереса к big data: поисковые запросы в Google в баллах<sup>2</sup>

С 2010-х гг. наблюдается рост применения цифровых технологий и в социологических исследованиях на уровне сбора и анализа данных как в количественной, так и в качественной стратегии исследования. Даже минимальное владение цифровыми технологиями и языками программирования (в том числе для анализа больших данных<sup>3</sup>) потенциально открывает для представителей социальных наук новые возможности в научной деятельности и в решении социально-прикладных задач, может приводить к экономии бюджета времени и принципиально иным объемам анализируемой информации. Однако зачастую исследовательские стратегии, базирующиеся на таких технологиях и инструментах, заимствуются из практики маркетинговых или бизнес-аналитических исследований, что ставит весьма серьезные вопросы относительно репрезентативности, валидности данных и интерпретации полученных результатов этими методами. Таким образом, вследствие своей масштабности и комплексности большие данные на сегодняшний день

<sup>2</sup> Динамика интереса рассчитана при помощи веб-аналитического инструментария Google Trends, в котором степень популярности поискового запроса рассчитывается не в частоте запросов, а в относительных числах – баллах, отражающих тренд популярности. Анализ проводился 27.02.2024 г.

<sup>3</sup> Большие данные не существуют вне цифровых технологий, которые в свою очередь не ограничиваются только большими данными.

не позволяют игнорировать их как возможный источник для сбора и анализа социологической информации, при этом из-за многомерной структуры их определения требуют крайне взвешенной, методологически прозрачной и тонкой аналитической работы. Вместе с тем в научной отечественной социологической литературе этим вопросам не уделяется должное внимание (за исключением нескольких исследовательских коллективов); переводы статей зарубежных коллег на русский язык фрагментарны и эпизодичны, хотя, бесспорно, значимы.

Целью статьи является предварительный анализ возможностей, ограничений и существующей к настоящему времени практики применения цифровых технологий (в том числе больших данных) в социологических исследованиях. Изменения, происходящие в объекте исследования, акцентирование на актуальных аспектах проблематики и возможности методик обуславливали выбор тех или иных путей сбора и анализа данных в изучении социальных проблем образования. Так, для изучения предпочтений и планов молодых людей при выборе образовательной и профессиональной траектории было достаточно анкетирования и рассмотрения распределений [Шубкин 1970], но изучение доступности качественного образования потребовало обращения к кластерному анализу [Константиновский 2008], новые ограничения и возможности обусловили сбор информации посредством Интернета [Кузнецов 2021], а перемены в ориентациях молодежи на уровни образования – обращение прежде всего к материалам статистики [Константиновский, Попова 2020]. Новый этап развития исследований потребовал для углубления анализа и интерпретации расширить круг применяемых методик: полнее (чем это делалось прежде) использовать методы сбора и анализа, связанные с цифровыми технологиями и большими данными.

## Большие данные и социология

Значимость цифровых технологий и больших данных для социологии как, собственно, и для любой другой социальной науки обусловлена прежде всего характеристиками больших данных, которые были рассмотрены в предыдущем разделе статьи. Говоря о перспективах использования больших данных в социальных науках, исследователи в той или иной степени обращают внимание как минимум на три ключевых момента – на сбор данных, их предварительную обработку/анализ и интерпретацию. На *рисунке 2* эти операции представлены в виде углов треугольника. Далее последовательно рассмотрим, с какими потенциальными трудностями и перспективами имеет дело социология в трех представленных углах треугольника.

### *Сбор данных*

Существует убеждение (в особенности широко распространенное среди представителей несоциальных наук), что новые данные способны представлять всю исследовательскую совокупность и потому претендуют на абсолютность, подлинность и конечность. Апологеты цифровых технологий и больших данных

полагают, что большие данные позволяют перейти от случайной выборки к использованию всех данных по принципу « $N = all$ ». Благодаря этому якобы появляется возможность уточнить старые и выявить новые связи. Кроме того, соглашаясь, что увеличение объема данных приводит к увеличению вероятности появления в них ошибок, они, вместе с тем, исходят из предпосылки, что большой и беспорядочный объем данных может привести к более точным результатам, чем немногочисленная случайная выборка (см. обзор: [Leitgöb et al. 2023]). По сути, сторонники этой позиции призывают отказаться от фундаментального положения, согласно которому качество статистических выводов зависит от свойств случайной выборки, а не от увеличения ее размера; то есть проблема репрезентативности – это только вопрос достаточного количества данных, и их большие объемы позволяют не замечать ошибки измерения и предвзятости [Maier-Шенбергер, Кукьер 2014].



Рисунок 2. Ключевые моменты использования больших данных в социальных науках

Однако подобные представления сталкиваются с многочисленной критикой ([Boyd, Crawford 2012; Giardullo 2016; Harford 2014; Lazer et al. 2014]), которая строится главным образом на обосновании противоположной позиции. Сам по себе размер не делает данные лучше, потому что в огромных объемах информации возникает множество разных смещений и искажений. В этой связи совершенно неясно, как большой размер может нивелировать такие проблемы, ведь со временем они накапливаются и превращаются в тотальный «шум», среди которого становятся совершенно неразличимы полезные для анализа «сигналы».

Помимо этого, необходимо учитывать ряд моментов, которые ставят под сомнение рассмотрение данных по принципу « $N = all$ », по крайней мере, на современном этапе развития и применения цифровых технологий и методов обработки больших массивов информации. Во-первых, большие данные если и могут выступать в качестве генеральной совокупности, то в лучшем случае относительно определенного сообщества пользователей – социальной сети (например, «ВКонтакте», «Одноклассники» и так далее), поисковой системы («Яндекс», «Google»), онлайн-платформы (кинотеатры, музыка, обучение), онлайн-приложений, корпоративных сайтов, видеохостингов и прочего. Хотя численность таких сообществ может достигать десятков миллионов человек, тем не менее они являются всего лишь фрагментарным отражением общества. То есть не совсем понятно, как

экстраполировать результаты, полученные в рамках определенного сообщества, на всю популяцию – на всю исследуемую совокупность, которая, по сути, является конститутивным признаком социологического исследования.

Во-вторых, всегда возникает вопрос, что это за данные, как и кем они получены, регистрируются ли одинаковым способом, возникают ли при этом искажения, не нарушается ли целостность при передаче. В.В. Волков с коллегами приводят следующий пример, который нередко встречается в повседневной жизни: «<...> если на проходной одного из заводов компании нет возможности обойти “вертушку” и каждое пересечение работником ворот фиксируется, то на соседнем заводе той же компании, данные с которого попадают в ту же базу данных, принято не прикладывать пропуск, выходя на перекур, и имеется техническая возможность обойти эту “вертушку”. В результате в базе мы увидим радикальные отличия по интенсивности движения через проходные двух заводов» [Волков и др. 2016, с. 51].

На примере *Twitter* Д. Бойд и К. Кроуфорд замечают схожую проблему, свойственную социальным сетям: «Мы также не можем предполагать, что учетные записи и пользователи эквивалентны. Некоторые пользователи имеют несколько учетных записей, а некоторые учетные записи используются несколькими людьми. Другие пользователи никогда не создают учетную запись и просто получают доступ к *Twitter* через Интернет. Также учетные записи могут представлять собой «ботов», которые автоматически создают контент без непосредственного участия человека» [Boyd, Crawford 2012, p. 669].

Т. Волбринг приводит другой пример в отношении мобильных устройств: «<...> геолокация мобильных телефонов с течением времени может очень хорошо описать индивидуальные модели перемещения для многих пользователей, но предоставляет бесполезную и вводящую в заблуждение информацию, если человек забыл смартфон дома или кому-то его одолжил. Огромные объемы информации не позволяют исследователям вручную искать подобные несоответствия и ошибки, что делает маловероятным их обнаружение и исправление» [Wolbring 2020, p. 65].

Помимо упомянутых существует большое количество других методических ограничений и подводных камней, возникающих при использовании цифровых технологий, больших данных и во время работы с ними (см. подробнее: [Olteanu et al. 2016; Mehrabi et al. 2021]). Ошибки и смещения могут быть связаны:

- с поведением пользователей (различия в поведении пользователей на разных платформах или в разных контекстах, при взаимодействии с контентом и другими пользователями);
- с производством контента (лексические, синтаксические, семантические и структурные различия в создаваемом контенте);
- с отсутствием или невозможностью оценки социально-демографических показателей, социально-экономического статуса пользователя, его идеологических/религиозных/политических пристрастий и тому подобного;
- с алгоритмическими и техническими особенностями социальных сетей, онлайн-платформ и устройств;
- с выбором источников данных или способа получения данных;
- с тем, каким образом формируется запрос об интересующей исследователя информации (по ключевым словам, на основе API и т. п.).

Такое положение дел требует внимательной работы с большими данными – понимания их специфических свойств и ограничений. Притом знание и учет



слабых мест важны и обязательны независимо от размеров массива информации. Если снова задаться вопросом о том, как все-таки рассматривать большие данные – как генеральную совокупность или как выборочную, то на основе вышесказанного следует выбрать, по всей видимости, второй вариант. В противном случае высока вероятность утратить необходимую бдительность и иметь дело со смещенной выборкой вместо «генеральной совокупности» (см., например: [Bail 2014; Wang et al. 2012]).

Кроме того, еще предстоит подробно изучить возможные решения и компенсации существующих затруднений, связанных с производством и накоплением больших данных. Пока же коротко перечислим некоторые стратегии, которые рекомендовали себя при сочетании разных методов получения информации и которые применяются или могут быть применены социальными исследователями в работе с большими данными.

1. Социологические исследования помогают прояснить контекст получения больших данных: например, качественные и/или количественные исследования позволяют продвинуться в понимании того, каким именно образом информация вносится в базы, какие критерии используются для ее регистрации, что она означает (см. [Волков и др. 2016]).
2. Большие данные могут быть дополнены социологическими исследованиями (или наоборот): речь идет о триангуляции разных источников данных и методов их сбора с целью повышения достоверности.
3. Большие данные служат основой для гипотез, которые затем проверяются при помощи социологических исследований.
4. Большие данные и социологические исследования дополняют друг друга, фокусируясь на специфических объектах исследования, которые оказываются вне поля зрения либо одного, либо другого способа получения информации: так, социология может уточнить социально-демографические характеристики респондентов, тогда как большие данные позволяют прояснить их включенность в социальные сети.

### *Обработка/анализ*

Многие устоявшиеся методы статистических расчетов не применимы к новому типу данных; речь идет в том числе о таких понятиях, как репрезентативность, стандартная ошибка и так далее. С одной стороны, это актуализирует различие случайных и неслучайных выборок, а с другой, ставит вопрос о способах оценки точности, достоверности, экстраполяции и т. п. в отношении больших массивов информации.

Обработка и анализ данных ранее невиданного объема, уровня детализации и сложности структуры требуют применения методов, которые могли бы адекватно учитывать их особенности. Новые методы основаны на автоматизированных технологиях анализа больших объемов информации; они расширяют и переопределяют возможности таких социологических методов, как сетевой и контент-анализ [McFarland et al. 2015; Amaturio, Aragona 2019]. С момента появления социограммы Я. Морено и теории поля К. Левина сетевой подход прошел большой путь в своем развитии. Сегодня он представлен как минимум тремя самостоятельными

направлениями – реляционной социологией, акторно-сетевой теорией и анализом социальных сетей [Мальцева 2017]. В последнем случае данный подход развивается главным образом с помощью математических и алгоритмических языков описания (статистического анализа и теории графов): так, эти языки позволяют сфокусироваться не только на моделировании сетей, но и на изучении сетевой динамики в реальном времени, а также перейти от исследований в малых группах к анализу социальных сетей в больших масштабах [Amaturo, Aragona 2019]. В свою очередь контент-анализ сегодня активно переосмысливается с позиций вычислительной лингвистики: в частности, данный подход демонстрирует вдохновляющие результаты благодаря развитию методов, которые основаны на построении моделей текстовых фрагментов на основе тематического моделирования [Коршунов, Гомзин 2012].

Методы машинного обучения в свою очередь тоже не стояли на месте, развивались вслед за развитием технологий, предлагая решения для разных типов данных – числового, текстового, видео, фото, аудио и прочего форматов. Более конкретными примерами могут служить методы обучения генеративных моделей; самообучения (*self-supervised learning*), извлечения информативных признаков (*feature extraction*) из данных; методы обучения с подкреплением (*reinforcement learning*) и другие. При этом следует отметить, что статистические способы анализа, которые задействованы в той или иной степени в машинном обучении, имеют давнюю традицию применения в социальных науках; среди них можно выделить линейное моделирование, кластеризацию k-средних, анализ главных компонент и др.

Алгоритмы обучения направлены как на построение прогнозных моделей, так и на описание систематических закономерностей (выявление скрытых структур в данных). Однако автоматизированные методы анализа все еще неспособны успешно решать проблему непрозрачности входных данных и обеспечивать значимые оценки неопределенности относительно построенных моделей [Grimmer et al. 2021].

Разумеется, алгоритмы могут совершенствоваться, обновляться и обучаться для получения более точных предсказаний. Однако остаются актуальными затруднения, связанные, с одной стороны, с ошибками, возникающими в результате производства, сбора, хранения и передачи данных, а с другой, с работой самих алгоритмов (подробнее см., например: [Mehrabi et al. 2021]). Кроме того, модели машинного обучения часто представляют собой «черные ящики», поскольку апеллируют к большому количеству переменных (которые могут исчисляться миллионами и характеризоваться множеством параметров) и выполняют сложные и ненаблюдаемые процедуры, приводящие к результатам, которым исследователи вынуждены доверять [Radford, Joseph 2020]. Непрозрачность алгоритмов затрудняет их анализ, а также методологическую и социальную критику [Diaz-Bone et al. 2020, p. 327]. Эти и подобные им ограничения лежат в основе часто встречаемого в литературе мнения, согласно которому методы машинного обучения все еще не способны выступать в качестве самостоятельного подхода к анализу данных, относящихся к социальной сфере.

В целом это означает, что для понимания и применения методов машинного обучения от исследователя в области социальных наук требуются как минимум знакомство, а по существу, знания и навыки прикладного программирования; то есть речь идет о приобщении к алгоритмической культуре статистического моделирования [Breiman 2001; Friedrich et al. 2022].



## Интерпретация

Осмысление проблем, связанных с двумя рассмотренными выше углами треугольника (рисунк 2, стр. 148), поставило вопрос о способах интерпретации новых данных. Развитие информационных и компьютерных технологий привело к появлению возрастающих и динамичных потоков информации, а также статистических пакетов, библиотек языков программирования, позволяющих с высокой скоростью их обрабатывать. Возникли петабайты данных, изначально не ориентированных на вычленение содержательного знания. В социальных науках интерес к большим данным проявился в тех областях, где особое значение имели данные, создаваемые без участия исследователя (например, социология науки) [Губа 2021]. Если в классическом (прежде всего количественном) социологическом исследовании сбору данных предшествуют выдвижение гипотез, разработка концептуального аппарата, выбор измерительного инструментария, то исследования, опирающиеся на большие данные, предлагают принципиально иную логику: «<...> данные – описание механизма их сбора – интерпретация переменных – гипотезы или вопросы – проверка» [Волков и др. 2016, с. 53]. Хотя в последнем случае представлена не устоявшаяся (и не единственная) логика, тем не менее она отражает суть эпистемологического сдвига, состоящего в том, что исследования больших данных в социологии (и в социальных науках в целом) должны или могут исходить из данных, а не из теории. В этой связи наметились два подхода, которые можно условно обозначить как «крайний» и «умеренный» эмпиризм.

Первый вариант характеризуется стремлением полностью упразднить необходимость обращения к теоретическим моделям и гипотезам [Anderson 2008]. «Крайний» эмпиризм опирается на предпосылку, согласно которой выявляемые взаимосвязи в больших данных уже сами по себе имеют значимость и содержательное значение, и потому достаточно разобраться в соответствующих статистических методах обработки и анализа данных, чтобы их читать и понимать [Китчин 2017, с. 122–123]. Апологет этого направления К. Андерсон в своем небольшом провокационном эссе красноречиво заявляет следующее: «Мы можем анализировать данные, не формулируя гипотезы о том, что в них содержится. Мы можем загрузить цифры в самые большие компьютерные кластеры, которые когда-либо существовали, и позволить статистическим алгоритмам обнаружить закономерности, которые не может выявить теория» (цитата по: [Китчин 2017, с. 120]). Призыв к отмене теорий, по сути, означает отмену наук, основанных на «старых» правилах. Андерсон продолжает: «Это мир, где огромные объемы данных и прикладная математика заменяют все остальные инструменты, которые могут быть использованы. Долой все теории человеческого поведения – от лингвистики до социологии. Забудьте о таксономии, онтологии и психологии. Кто знает, почему люди делают то, что делают? Главное, что они это делают, и мы можем отслеживать и измерять это с беспрецедентной точностью. Если данных достаточно, цифры говорят сами за себя» [Anderson 2008]. Общим для этого подхода является построение исследования и науки на индуктивном методе. Иными словами, при работе с большими данными предлагается рассматривать треугольник (рисунк 2, стр. 148) в усеченном виде без апелляции к интерпретации; редуцировать эту операцию к простому описанию данных – корреляционному анализу – без установления причинно-следственных связей.

Реакция со стороны научного сообщества не заставила себя долго ждать [Kitchin 2017; Wolbring 2020]. Р. Китчин выдвинул несколько контраргументов, среди которых отметим два наиболее важных. Первый состоит в том, что данные не могут быть получены независимо от теории. В основе вычислительных технологий, методов и алгоритмов анализа уже лежат логика, теоретические допущения, предшествующая научная работа и опыт, обеспечивающие их валидность и достоверность. Вторым аргументом сводится к тому, что данные не способны «говорить сами за себя» [Kitchin 2017, с. 123–124]. На них оказывают влияние разные «картины мира» исследователя – научная, корпоративная, здравого смысла или какая-либо еще, и это во многом определяет, что он может «услышать» или «увидеть». Исследование в той или иной степени теоретически нагружено, и важно понимать, в какой именно системе координат оно проводится, чтобы, например, лучше оценить его объяснительные возможности или заметить случайную природу установленной корреляции между переменными. В этом смысле «крайний» эмпиризм предлагает, по сути, экспансию одного способа теоретического описания – математического, то есть языка алгоритмов, статистики и вероятностей. Также он отражает претензию заменить этим языком все другие возможные языки описания «от лингвистики до социологии».

Добавим, что «крайний» эмпиризм вряд ли сможет стать серьезным вызовом для современной социологической науки по ряду причин: во-первых, в силу все еще существующих затруднений, связанных с природой больших данных, с их сбором, обработкой и анализом; во-вторых, в социологии объяснительный потенциал апеллирует к теориям, которые имеют сильную связь с философскими основаниями; и, наконец, в-третьих, благодаря философии науки хорошо известно различие эмпиризма и априоризма – противостояние индуктивного и дедуктивного способов получения знания, которое уже неоднократно воплощалось в социологической науке в виде спора «языка» и «метода», «теоретиков» и «эмпириков», а также «физиков» и «лириков».

«Умеренный» эмпиризм стремится сгладить острые углы рассмотренного выше подхода. Он сочетает индуктивные и дедуктивные аспекты, ориентируясь на существующую научную методологию [Kitchin 2017; McFarland et al. 2015]. При этом индуктивное объяснение в отличие от «крайнего» эмпиризма не становится конечной целью, а создает лишь условия для последующей теоретической работы и проверки гипотез. Более того, теория не просто учитывает индуктивные инсайды, но также абдуктивно и дедуктивно направляет поиск и открытие новых знаний. Отношение между индукцией, абдукцией и дедукцией отражается в принципе интеракции, который сводится к «сшиванию» эмпирических и теоретических реальностей, когда исследователь постоянно обращается то к одному, то к другому полюсу социологического исследования. Благодаря индукции возникают определенные предположения, догадки, служащие основанием для выдвижения гипотез; абдукция позволяет сосредоточиться не на всех, а на конкретных взаимосвязях и гипотезах, которые, по мнению эксперта, могут выглядеть более разумными и обоснованными в свете уже известных научных достижений в конкретной области; эти предположения и допущения дедуктивно проверяются [Kitchin 2017, с. 128], и затем подобная проверка подкрепляется повторным обращением к массиву больших данных. Такое «реверсивное движение» преследует цель сохранить объяснительный потенциал исследования и не ограничивать его поиском только паттернов или закономерностей [Halavais 2015].

Несмотря на то, что подобное взаимоотношение между индукцией, абдукцией и дедукцией не является новой идеей и практикой в социологическом исследовании, тем не менее сейчас, как справедливо замечает К.С. Губа, оно рассматривается как легитимный вариант будущей эмпирически ориентированной социальной науки [Губа 2018].

Существует еще одна альтернативная позиция, которая подразумевает, что появление больших данных и развитие методов машинного обучения делают теорию не менее, а, наоборот, более актуальной. Эту позицию можно выразить несколько иначе: перспективы современных социальных исследований, основанных на больших данных и применении машинного обучения, становятся многообещающими именно и только благодаря применению и развитию теорий и методологии исследования [Radford, Joseph 2020; Wolbring 2020]. В данном случае вектор построения социологического исследования меняется не от данных к теории, а от теории к данным, то есть отправной точкой исследования являются не столько данные, сколько теоретические представления. В основе такого разворота лежит аксиоматическое предположение, согласно которому большие массивы информации, несмотря на их особенности, не отличаются по способу своего изучения от любого другого эмпирического материала. Аргументы сторонников дедуктивного подхода сводятся к тому, что данная логика позволяет уменьшить количество ложных «открытий» и изначально заложить в исследование концептуальные представления, которые позволяют поместить изучаемый феномен в определенную категориальную сетку, тем самым создав условия для интерпретации и объяснения. Дж. Рэдфорд и К. Джозеф предупреждали, что обращение только к «компьютерным наукам» (методам машинного обучения) для работы с социальными данными может привести в своем наихудшем варианте к «<...> псевдонауке, когда неверно примененные алгоритмы используются для принятия дискриминационных решений и для выдвижения необоснованных социально-научных утверждений» [Radford, Joseph 2020].

Сопряжение больших данных, машинного обучения и социальных наук открывает многообещающие перспективы в науке. Тем не менее все еще мало внимания уделяется моделям машинного обучения, которые основаны на более глубоком понимании общества [Radford, Joseph 2020]. Это позволяет развивать техническую, математическую и алгоритмическую стороны синтеза больших данных, машинного обучения и социологии без должного внимания к достижениям в области социальных наук в целом. Можно заключить, что «умеренный» эмпиризм и априоризм способны стать отправными точками для плодотворного обсуждения методологии, эпистемологии и принципов получения социологических знаний на основе больших данных.

## Заключение

По масштабу, объему и сложности современный этап технологического развития не имеет аналогов во всем предыдущем опыте человечества. Процессы цифровизации значимо трансформировали многие сферы социальной жизни – медицину, образование, культуру, рынок труда, экономику и другие. По каким-то направлениям

внедрение цифровых технологий проходило в плавном, «сглаженном» режиме, а в каких-то сферах пришлось экстренно осваивать новые – цифровые – технологии, как это случилось в образовании в период пандемии.

Сфера социологических исследований, как было показано, не является исключением: цифровые технологии сбора и анализа данных прочно укореняются в практике исследований социальных наук. При анализе этих процессов отмечают, как инновационная природа применяемых цифровых технологий, которая меняет исследование, создавая возможность формирования новых подходов в науке [Китчин 2017, с. 141], так и необходимость пересмотра онтологических, эпистемологических и методологических принципов социологии [Щекотин 2020, с. 54].

В исследовательском сообществе активно обсуждаются вопросы, как изучать подобные потоки информации. Широко распространено мнение, что большие данные по причине своей специфичности предполагают работу над методологическими изменениями в процедурах их обработки и анализа. Эти изменения, в частности, обусловлены тем, что социолог обычно не принимает участия в создании больших данных, не собирает их по заранее подготовленным кабинетным процедурам и теоретическим допущениям, а имеет дело с уже готовыми массивами информации. Как правило, исследователь не только не влияет на инструмент сбора больших данных, но и не контролирует его в действии [Волков и др. 2016, с. 51]. За производство подобных массивов данных отвечают цифровые технологии, устройства и интерфейсы, которые составляют сегодня привычный образ жизни [McFarland et al. 2015]. Безусловно, это обеспечивает позитивную возможность избежать в некоторых случаях «искусственного» характера получения информации, которая составляет отличительную черту опросных методик. Например, миновать ситуацию, когда респондент вынужден отвечать на вопросы, о которых он никогда не задумывался, и в этом заключается привлекательность новых массивов информации. Но в то же время обращение к ним – это и вызов для социальных исследователей; последнее сопряжено с особенностями извлечения полезного знания из больших данных и в целом с их применением в социологии.

Одна из основных задач, которая, на наш взгляд, стоит сегодня перед социологами в практике применения цифровых технологий и больших данных в социологических исследованиях, заключается в обеспечении методологической прозрачности от сбора данных до их анализа и ответственном выборе алгоритмов машинного обучения применительно к конкретным исследовательским задачам<sup>4</sup>. В больших данных есть все, но абсолютно не все нам нужно для анализа: машина не понимает специфики данных, а для решения социально-прикладных задач критично понимать именно контекст сбора данных, их природу; социологическими методами обеспечивать реконструкцию методики сбора данных и их оценку. Более того, работа с большими данными – это циклический

<sup>4</sup> Приведем здесь, на наш взгляд, значимое высказывание Х. Уоллак, старшего научного сотрудника Microsoft Research, в котором отражена суть сопряжения больших данных, машинного обучения и социологии: «Очевидно, что машинное обучение невероятно полезно, и, в частности, полезно для социальных наук. Но мы должны относиться к машинному обучению в социальных науках совершенно иначе, чем, скажем, при распознавании рукописного текста или игре в шахматы. Мы не можем просто применять методы машинного обучения по принципу «черного ящика», как будто вычислительная социальная наука (*computational social science*) – это просто информатика плюс социальные данные. <...> Мы должны уделять приоритетное внимание интерпретируемости – даже в контексте прогнозирования. Мы должны проводить тщательный и детальный анализ ошибок. <...> Но, что наиболее важно, мы должны работать совместно с социальными учеными, чтобы понимать этические последствия наших решений по моделированию» [Wallach 2018, p. 44].

процесс, так как итерации по очистке, обработке, анализу данных могут повторяться неоднократно. Даже при помощи самых продвинутых методов визуализации мы не видим массив целиком, работа ведется по настраиваемым исследователем срезам. В процессе исследования интерпретируются не сами данные, а те эмпирические закономерности, которые в этих данных обнаруживаются. Если переходить на язык метафор, под большими данными с одинаковой вероятностью можно получить как «черный ящик» (*black box*), который предстоит расшифровать и интерпретировать, или «пустышку» (*nothing box*), непригодную для анализа. И третий момент, который необходимо учитывать именно в социологических исследованиях: большие данные не являются самодостаточными данными, это один из видов данных. Поэтому ответственность социолога – использовать метод триангуляции данных для обеспечения прозрачности, взвешенности и полноты интерпретации результатов, что в статьях о больших данных сегодня почти не встречается.

### Список источников

- Волков В.В., Скугаревский Д.А., Титаев К.Д. (2016) Проблемы и перспективы исследований на основе Big Data (на примере социологии права) // Социологические исследования. № 1. С. 48–58 // <https://www.socis.isras.ru/article/6015>, дата обращения 11.03.2024.
- Губа К.С. (2018) Большие данные в социологии: новые данные, новая социология? // Социологическое обозрение. Т. 18. № 1. С. 213–236 // <https://sociologica.hse.ru/2018-17-1/217766689.html>, дата обращения 11.03.2024.
- Губа К.С. (2021) Большие данные в исследовании науки: новое исследовательское поле // Социологические исследования. № 6. С. 24–33. DOI: 10.31857/S013216250013878-8
- Гурвиц Дж., Ньюджент А., Халпер Ф., Кауфман М. (2015) Просто о больших данных. М.: Эксмо.
- Китчин Р. (2017) Большие данные, новые эпистемологии и смена парадигм // Социология: методология, методы, математическое моделирование. № 44. С. 111–152 // [https://www.isras.ru/index.php?page\\_id=2384&id=5609&l=1](https://www.isras.ru/index.php?page_id=2384&id=5609&l=1), дата обращения 11.03.2024.
- Константиновский Д.Л. (2008) Неравенство и образование. Опыт социологических исследований жизненного старта российской молодежи (1960-е годы – начало 2000-х). М.: ЦСО.
- Константиновский Д.Л., Попова Е.С. (2020) Среднее vs высшее // Мир России. Т. 29. № 2. С. 6–26. DOI: 10.17323/1811-038X-2020-29-2-6-26
- Коршунов А., Гомзин А. (2012) Тематическое моделирование текстов на естественном языке. Труды Института системного программирования РАН. Т. 23. DOI: 10.15514/ISPRAS-2012-23-13
- Кузнецов И.С. (2021) Детерминанты доверия в высшем образовании // Высшее образование в России. Т. 30. № 12. С. 9–31. DOI: 10.31992/0869-3617-2021-30-12-9-31
- Майер-Шенбергер В., Кукьер К. (2014) Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М.: Манн, Иванов и Фербер.
- Мальцева Д.В. (2017) Сетевой подход в социологии: генезис идей и применение. Новосибирск: НГТУ.
- Шубкин В.Н. (1970) Социологические опыты. Методологические вопросы социальных исследований. М.: Мысль.
- Щекотин Е.В. (2020) Цифровые технологии в социальных науках: предмет и метод цифровой социологии // Социология и право. № 1. С. 49–59. DOI: 10.35854/2219-6242-2020-1-49-59



- Amaturo E., Aragona B. (2019) Methods for Big Data in Social Sciences // *Mathematical Population Studies*, vol. 26, no 2, pp. 65–68. DOI: 10.1080/08898480.2019.1597577
- Anderson Ch. (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete // *WIRED*, June 23, 2008 // <https://www.wired.com/2008/06/pb-theory/>, дата обращения 11.03.2024.
- Bail S.A. (2014) The Cultural Environment: Measuring Culture with Big Data // *Theory and Society*, vol. 43, no 3, pp. 465–482. DOI: 10.1007/s11186-014-9216-5
- Boyd D., Crawford K. (2012) Critical Questions for Big Data // *Information, Communication & Society*, vol. 15, no 5, pp. 662–679. DOI: 10.1080/1369118X.2012.678878
- Breiman L. (2001) Statistical Modeling: The Two Cultures // *Statistical Science*, vol. 16, no 3, pp. 199–231. DOI: 10.1214/ss/1009213726
- Diaz-Bone R., Horvath K., Cappel V. (2020) Social Research in Times of Big Data. The Challenges of New Data Worlds and the Need for a Sociology of Social Research // *Historical Social Research / Historische Sozialforschung*, vol. 45, no 3, pp. 314–341 // <https://www.jstor.org/stable/26918415>, дата обращения 11.03.2024.
- Friedrich S. et al. (2022) Is There a Role for Statistics in Artificial Intelligence? // *Advances in Data Analysis and Classification*, vol. 16, no 4, pp. 823–846. DOI: 10.1007/s11634-021-00455-6
- Giardullo P. (2016) Does ‘Bigger’ Mean ‘Better’? Pitfalls and Shortcuts Associated with Big Data for Social Research // *Quality & Quantity: International Journal of Methodology*, vol. 50, no 2, pp. 529–547 // <https://ideas.repec.org/a/spr/qualqt/v50y2016i2p529-547.html>, дата обращения 11.03.2024.
- Grimmer J., Roberts M.E., Stewart B.M. (2021) Machine Learning for Social Science: An Agnostic Approach // *Annual Review of Political Science*, vol. 24, pp. 395–419 // <https://www.annualreviews.org/doi/10.1146/annurev-polisci-053119-015921>, дата обращения 11.03.2024.
- Halavais A. (2015) Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods // *Information, Communication & Society*, vol. 18, no 5, pp. 583–594. DOI: 10.1080/1369118X.2015.1008543
- Harford T. (2014) Big Data: A Big Mistake? // *Significance*, vol. 11, pp. 14–19. DOI: 10.1111/j.1740-9713.2014.00778.x
- Lazer D., Kennedy R., King G., Vespignani A. (2014) The Parable of the Google Flu: Traps in Big Data Analysis // *Science*, vol. 343, no 6176, pp. 1203–1205. DOI: 10.1126/science.1248506
- Leitgöb H., Prandner D., Wolbring T. (2023) Editorial: Big Data and Machine Learning in Sociology // *Frontiers in Sociology*, vol. 8, art. 1173155. DOI: 10.3389/fsoc.2023.1173155
- McFarland D.A., Lewis K., Goldberg A. (2015) Sociology in the Era of Big Data: The Ascent of Forensic Social Science // *American Sociologist*, vol. 47, no 1, pp. 12–35 // <https://www.gsb.stanford.edu/faculty-research/publications/sociology-era-big-data-ascent-forensic-social-science>, дата обращения 11.03.2024.
- Mehrabi N., Morstatter F., Saxena N., Lerman K., Galstyan A. (2021) A Survey on Bias and Fairness in Machine Learning // *ACM Computing Surveys*, vol. 54, no 6, pp. 1–35. DOI: 10.1145/3457607
- Olteanu A., Castillo C., Diaz F., Kiciman E. (2016) Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries // *SSRN Scholarly Paper ID 2886526*, Rochester, New York: Social Science Research Network. DOI: 10.2139/ssrn.2886526
- Radford J., Joseph K. (2020) Theory In, Theory Out: The Uses of Social Theory in Machine Learning for Social Science // *Frontiers in Big Data*, vol. 3, art. 18. DOI: 10.3389/fdata.2020.00018
- Wallach H. (2018) Computational Social Science ≠ Computer Science + Social Data // *Communications of the ACM*, vol. 61, no 3, pp. 42–44. DOI: 10.1145/3132698
- Wang D.J., Shi X., McFarland D.A., Leskovec J. (2012) Measurement Error in Social Network Data: A Re-classification // *Social Networks*, vol. 34, no 4, pp. 396–409 // <https://cs.stanford.edu/people/jure/pubs/error-socnets12.pdf>, дата обращения 11.03.2024.



Wolbring T. (2020) The Digital Revolution in the Social Sciences: Five Theses about Big Data and Other Recent Methodological Innovations from an Analytical Sociologist // Sociology of the Digital – Digital Sociology (eds. Maasen S., Passoth J.-H.), Baden-Baden: Nomos, pp. 60–72.

---

## Digital Technologies and Big Data in Sociological Research: Concept, Methodology, Opportunities

---

D.L. KONSTANTINOVSKIY\*, E.S. POPOVA\*\*, I.S. KUZNETSOV\*\*\*,  
R.S. KUZNETSOV\*\*\*\*

\***David L. Konstantinovskiy** – DSc in Sociology, Head of Department of Sociology of Education, Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russian Federation; scan21@mail.ru, <https://orcid.org/0000-0003-3316-0644>

\*\***Ekaterina S. Popova** – PhD in Sociology, Leading Researcher, Department of Sociology of Education, Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russian Federation; espopova@isras.ru, <https://orcid.org/0000-0002-9808-3152>

\*\*\***Igor S. Kuznetsov** – PhD in Sociology, Researcher, Department of Sociology of Education, Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russian Federation; kuznetsov.igor@gmail.com, <https://orcid.org/0000-0002-4477-5233>

\*\*\*\***Roman S. Kuznetsov** – Junior Researcher, Department of Sociology of Education, Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russian Federation; roman@c-discurs.ru, <https://orcid.org/0000-0003-3063-9465>

**Citation:** Konstantinovskiy D.L., Popova E.S., Kuznetsov I.S., Kuznetsov R.S. (2025) Digital Technologies and Big Data in Sociological Research: Concept, Methodology, Opportunities. *Mir Rossii*, vol. 34, no 1, pp. 144–160 (in Russian). DOI: 10.17323/1811-038X-2025-34-1-144-160

### Abstract

*Digitalization has significantly transformed numerous spheres of social life. While the implementation of digital technologies has been smooth in some areas, in others it has required urgent adaptation. The field of sociological research is no exception: digital technologies for data collection and analysis have become deeply integrated into social science research practices. The use of digital tools and big data analysis in sociology offers a significant resource for breakthrough insights into social reality, but it also poses substantial risks when it comes to explaining this reality and developing practical recommendations for addressing social issues.*

*Before engaging with digital technologies, tools, and big data, it is essential to clarify their theoretical and methodological nature as well as their practical potential from a social science perspective. This article analyzes the opportunities, limitations, and initial lessons from current practices in applying digital technologies, including big data, to sociological research.*

---

The authors express their gratitude to P.M. Kozyreva for support in the development of the direction, E.I. Grigorieva for help and valuable advice, M.S. Kosolapov for deep questions on the concept and methodology.

The article was received in March 2024.

The article explores the unique features of collecting and analyzing such data and examines various methods of interpreting it.

Particular attention is given to the intersection of social science, big data, and machine learning. It is noted that little attention has been paid to the application of machine learning models in sociology. The need to develop the mathematical and algorithmic aspects of this synthesis is highlighted. The article suggests that a balanced approach between empiricism and apriorism can serve as a starting point for a productive discussion on the methodology, epistemology, and principles of sociological knowledge derived from big data. The role and responsibility of sociologists in solving practical social problems through big data analysis are also discussed, emphasizing the need for methodological rigor. Finally, the necessity of applying data triangulation is underscored.

**Keywords:** sociological research, sociological methods, methodology, data analysis, digitalization, digital technologies, machine learning methods, big data

## References

- Amaturo E., Aragona B. (2019) Methods for Big Data in Social Sciences. *Mathematical Population Studies*, vol. 26, no 2, pp. 65–68. DOI: 10.1080/08898480.2019.1597577
- Anderson Ch. (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *WIRED*, June 23, 2008. Available at: <https://www.wired.com/2008/06/pb-theory/>, accessed 11.03.2024.
- Bail C.A. (2014) The Cultural Environment: Measuring Culture with Big Data. *Theory and Society*, vol. 43, no 3, pp. 465–482. DOI: 10.1007/s11186-014-9216-5
- Boyd D., Crawford K. (2012) Critical Questions for Big Data. *Information, Communication & Society*, vol. 15, no 5, pp. 662–679. DOI: 10.1080/1369118X.2012.678878
- Breiman L. (2001) Statistical Modeling: The Two Cultures. *Statistical Science*, vol. 16, no 3, pp. 199–231. DOI: 10.1214/ss/1009213726
- Diaz-Bone R., Horvath K., Cappel V. (2020) Social Research in Times of Big Data. The Challenges of New Data Worlds and the Need for a Sociology of Social Research. *Historical Social Research / Historische Sozialforschung*, vol. 45, no 3, pp. 314–341. Available at: <https://www.jstor.org/stable/26918415>, accessed 11.03.2024.
- Friedrich S. et al. (2022) Is There a Role for Statistics in Artificial Intelligence? *Advances in Data Analysis and Classification*, vol. 16, no 4, pp. 823–846. DOI: 10.1007/s11634-021-00455-6
- Giardullo P. (2016) Does ‘Bigger’ Mean ‘Better’? Pitfalls and Shortcuts Associated with Big Data for Social Research. *Quality & Quantity: International Journal of Methodology*, vol. 50, no 2, pp. 529–547. Available at: <https://ideas.repec.org/a/spr/qualqt/v50y2016i2p529-547.html>, accessed 11.03.2024.
- Grimmer J., Roberts M.E., Stewart B.M. (2021) Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science*, vol. 24, pp. 395–419. Available at: <https://www.annualreviews.org/doi/10.1146/annurev-polisci-053119-015921>, accessed 11.03.2024.
- Guba K.S. (2018) Big Data in Sociology: New Data, New Sociology? *The Russian Sociological Review*, vol. 17, no 1, pp. 213–236. Available at: <https://sociologica.hse.ru/2018-17-1/217766689.html>, accessed 11.03.2024 (in Russian).
- Guba K.S. (2021) Big Data in Studies of Science: New Research Field. *Sociological Studies*, no 6, pp. 24–33 (in Russian). DOI: 10.31857/S013216250013878-8
- Halavais A. (2015) Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods. *Information, Communication & Society*, vol. 18, no 5, pp. 583–594. DOI: 10.1080/1369118X.2015.1008543

- Harford T. (2014) Big Data: A Big Mistake? *Significance*, vol. 11, pp. 14–19. DOI: 10.1111/j.1740-9713.2014.00778.x
- Hurwitz J., Nugent A., Halper F., Kaufman M. (2013) *Big Data for Dummies*, Moscow: Eksmo (in Russian).
- Kitchin R. (2017) Big Data, New Epistemologies and Paradigm Shifts. *Sociology: Methodology, Methods, Mathematical Modeling*, no 44, pp. 111–152. Available at: [https://www.isras.ru/index.php?page\\_id=2384&id=5609&l=,](https://www.isras.ru/index.php?page_id=2384&id=5609&l=,) accessed 11.03.2024 (in Russian).
- Konstantinovskiy D.L. (2008) *Inequality and Education. Attempt of Sociological Research on the Life Starts of the Russian Youth (1960<sup>th</sup> – Beginning of 2000<sup>th</sup>)*, Moscow: SFC (in Russian).
- Konstantinovskiy D.L., Popova E.S. (2020) Vocational vs Higher Education. *Mir Rossii*, vol. 29, no 2, pp. 6–26 (in Russian). DOI: 10.17323/1811-038X-2020-29-2-6-26
- Korshunov A., Gomzin A. (2012) Topic Modeling in Natural Language Texts. *Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS)*, vol. 23 (in Russian). DOI: 10.15514/ISPRAS-2012-23-13
- Kuznetsov I.S. (2021) Determinants of Trust in Higher Education. *Higher Education in Russia*, vol. 30, no 12, pp. 9–31 (in Russian). DOI: 10.31992/0869-3617-2021-30-12-9-31
- Lazer D., Kennedy R., King G., Vespignani A. (2014) The Parable of the Google Flu: Traps in Big Data Analysis. *Science*, vol. 343, no 6176, pp. 1203–1205. DOI: 10.1126/science.1248506
- Leitgöb H., Prandner D., Wolbring T. (2023) Editorial: Big Data and Machine Learning in Sociology. *Frontiers in Sociology*, vol. 8, art. 1173155. DOI: 10.3389/fsoc.2023.1173155
- Maltseva D.V. (2017) *The Network Approach in Sociology: Genesis of Ideas and Applications*, Novosibirsk: NSTU Publisher (in Russian).
- Mayer-Schönberger V., Cukier K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Moscow: Mann, Ivanov, Ferber (in Russian).
- McFarland D.A., Lewis K., Goldberg A. (2015) Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *American Sociologist*, vol. 47, no 1, pp. 12–35. Available at: <https://www.gsb.stanford.edu/faculty-research/publications/sociology-era-big-data-ascent-forensic-social-science>, accessed 11.03.2024.
- Mehrabi N., Morstatter F., Saxena N., Lerman K., Galstyan A. (2021) A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, vol. 54, no 6, pp. 1–35. DOI: 10.1145/3457607
- Olteanu A., Castillo C., Diaz F., Kiciman E. (2016) Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *SSRN Scholarly Paper ID 2886526*, Rochester, New York: Social Science Research Network. DOI: 10.2139/ssrn.2886526
- Radford J., Joseph K. (2020) Theory In, Theory Out: The Uses of Social Theory in Machine Learning for Social Science. *Frontiers in Big Data*, vol. 3, art. 18. DOI: 10.3389/fdata.2020.00018
- Shchekotin E.V. (2020) Digital Technologies in Social Sciences: Subject and Method of Digital Sociology. *Sociology and Law*, no 1, pp. 49–59 (in Russian). DOI: 10.35854/2219-6242-2020-1-49-59
- Shubkin V.N. (1970) *Sociological Experiments. Methodological Issues of Social Research*, Moscow: Mysl' (in Russian).
- Volkov V.V., Skougarevskiy D.A., Titaev K.D. (2016) Problems and Prospects for Studies Based on Big Data (the Case of Sociology of Law). *Sociological Studies*, no 1, pp. 48–58. Available at: <https://www.socis.isras.ru/article/6015>, accessed 11.03.2024 (in Russian).
- Wallach H. (2018) Computational Social Science ≠ Computer Science + Social Data. *Communications of the ACM*, vol. 61, no 3, pp. 42–44. DOI: 10.1145/3132698
- Wang D.J., Shi X., McFarland D.A., Leskovec J. (2012) Measurement Error in Social Network Data: A Re-classification. *Social Networks*, vol. 34, no 4, pp. 396–409. Available at: <https://cs.stanford.edu/people/jure/pubs/error-socnets12.pdf>, accessed 11.03.2024.
- Wolbring T. (2020) The Digital Revolution in the Social Sciences: Five Theses about Big Data and Other Recent Methodological Innovations from an Analytical Sociologist. *Sociology of the Digital – Digital Sociology* (eds. Maasen S., Passoth J.-H.), Baden-Baden: Nomos, pp. 60–72.