СОЦИОЛОГИЯ ИКТ

УДК 316.77

2020.02.009. ПАВЛЮЧЕНКО А.А. РЕЦЕНЗИЯ НА КН.: СТИВЕНС-ДАВИДОВИЦ С. ВСЕ ЛГУТ: ПОИСКОВИКИ, BIG DATA И ИНТЕРНЕТ ЗНАЮТ О ВАС ВСЁ / пер. с англ. Л.И. Степановой. – М.: Эксмо, 2018. – 384 с.

PAVLUCHENKO A.A. Book review: Stephens-Davidowitz S. Everybody lies: Big data, new data and what the Internet can tell us about who we really are. DOI: 10.31249/rsoc/2020.02.04

Аннотация. Книга бывшего аналитика Google С. Стивенса-Давидовица является популярным введением в мир «больших данных». Автор рецензии отмечает, что эффективное применение анализа доступной в Интернете информации, объединенной в массив «больших данных», позволяет проводить максимально достоверные исследования в сфере социогуманитарных и естественных наук.

Abstract. The book written by former Google data analyst S. Stephens-Davidowitz is a popular introduction to the world of Big Data. The author of the book review points out that effective use of the analysis of information available on the Internet, combined into an array of «big data», allows for utterly reliable research in the field of social sciences and humanities as well as natural sciences.

Ключевые слова: интервью; методология; Big Data; большие данные; США; Д. Трамп; Б. Обама.

Keywords: interview; methodology; Big Data; USA; D. Trump; B. Obama.

Люди склонны преувеличивать и недоговаривать; исследования недостаточно репрезентативны; опросы превратились в средство политической борьбы и повод для генерации определен-

ных новостей, компании манипулируют статистикой в стремлении упрочить свою репутацию на рынке, – таким образом, все лгут. В подобных условиях, несмотря на избыточное количество находящихся в открытом доступе цифровых данных, прогнозирование и анализ общественно-политических процессов парадоксальным образом осложняются.

Книга Стивенса-Давидовица – это одна из первых попыток систематизировать и оценить сложившиеся на сегодняшний день способы анализа массивов больших данных в гуманитарных и естественных науках, а также наметить основной вектор развития данного направления. По словам самого автора, эта работа призвана стать новой ступенью «фрикономики» Стивена Левитта [Левитт, Дабнер, 2007] и заразить молодое поколение исследователей идеей о возможности рассматривать анализ «больших данных» в качестве ключа к пониманию действительного устройства мира.

Основное различие между исследованиями Стивенса-Давидовица и Левитта заключается в том, что последний работал в 1990-х годах, когда данных, доступных для анализа, было не так много. По этой причине он зачастую игнорировал те области, которые ранее не описывались через количественные показатели. Однако в условиях современного мира сложно говорить о дефиците информации. Количество доступных для анализа тем не ограничено, а малозатратность анализа «больших данных» позволяет любому исследователю (при наличии навыка работы со статистикой и творческого подхода) задействовать этот инструмент в своих изысканиях.

С развитием Интернета возможности научного сообщества расширились. Использование доступной в Сети информации, объединенной в массив «больших данных» для построения прогнозов и проведения исследований, было признано одним из наиболее перспективных направлений в науке.

Однако «выделить подлинный сигнал в переполненной шумом Вселенной данных» оказалось не так просто. Одним из первых принципиально новую методологию оценки и подбора информации предложил американский социолог Нейт Сильвер [Silver, 2012]. «Рациональный прогрессивист», автор блога *Five-ThirtyEight* на сайте «New York Times», верно определивший победителя президентских выборов в США в 2012 г., стал олицетворе-

нием предсказательного тренда американской политической социологии.

В данном контексте книга бывшего аналитика Google Сета Стивенса-Давидовица, рассматриваемая в настоящей рецензии, является своего рода популярным введением в мир «больших данных». Под *Big Data* автор понимает текст и ссылки; фото и видео; поисковые запросы в Интернете. В совокупности они формируют не только персональные ячейки знаний о каждом конкретном пользователе, но представляют мощнейший инструмент для анализа [Стивенс-Давидовиц, 2018, с. 20]. С точки зрения Стивенса-Давидовица, люди не доверяют свои секреты интервьюерам социологических служб, однако раскрываются через собственные поисковые запросы, что позволяет эффективно использовать инструменты того же Google для отслеживания общественных настроений. Так, после предварительных республиканских выборов 2016 г. многие эксперты пришли к выводу, что у Дональда Трампа нет никаких шансов победить, так как во время предвыборной кампании он успел оскорбить почти все возможные меньшинства. Результаты опросов демонстрировали, что подобную риторику одобряет минимальное число американцев [Стивенс-Давидовиц, 2018, с. 12].

В ходе исследований, проводившихся в период президентства Барака Обамы, социологи раз за разом приходили к выводу, что раса не является существенным фактором при выборе кандидата. В академических кругах США утвердилось мнение, что расизм присущ только небольшому проценту маргиналов; некоторые исследователи даже утверждали, что в стране уже сформировалось пострасовое общество [Стивенс-Давидовиц, 2018, с. 13]. Однако с победой Д. Трампа эти иллюзии быстро развеялись.

В ходе собственного анализа поисковых запросов Google Стивенс-Давидовиц пришел к ряду значимых выводов. Во-первых, он обнаружил, что большая часть американцев скрывает свой расизм [Стивенс-Давидовиц, 2018, с. 19]. Во-вторых, автор выявил расхождение между устоявшимися представлениями о специфике локализации современного расизма и реальной географией его распространения. Согласно стереотипному восприятию, расизм распространен преимущественно среди республиканцев на Юге США. Однако, по мнению Стивенса-Давидовица, правильное было

бы противопоставлять не Север и Юг, а Запад и Восток. Исследование с использованием Google Trends показало, что наибольшее количество интернет-запросов расистского толка приходилось на IP-адреса, зарегистрированные в штатах Нью-Йорк, Пенсильвания, Западный и Восточный Огайо, Мичиган и Иллинойс [Стивенс-Давидовиц, 2018, с. 19]. Таким образом, можно прийти к выводу, что степень расизма жителей Юга – не больше, чем жителей Севера, просто южане откровеннее в выражении своих взглядов [Стивенс-Давидовиц, 2018, с. 20].

Это не единственные искажения, которые были обнаружены автором. Больше половины граждан, которые в итоге не голосуют, во время опросов говорят о своей готовности прийти на выборы. Изучение таких поисковых запросов Google, как «где голосовать?» и «как голосовать?» за неделю перед выборами, куда вернее покажет реальную явку избирателей, настаивает Стивенс-Давидовиц [Стивенс-Давидовиц, 2018, с. 21].

С его точки зрения, анализ «больших данных» позволяет проводить исследования, отбрасывая когнитивные искажения опрашиваемых, и получать информацию в тот момент, когда респондент не подвергается какому-либо социальному давлению.

Основываясь на этом принципе, автор делит доступную для анализа информацию на «цифровую правду» и «цифровую ложь». К первой относятся поисковые запросы, просмотры, клики и выделения текста, а ко второй – посты и лайки в социальных сетях, а также профили на сайтах знакомств [Стивенс-Давидовиц, 2018, с. 192].

По мнению автора, оптимизация методов анализа и сбора «больших данных» позволит социальным наукам приобрести точность, в наличии которой им часто отказывают представители точных и естественных наук [Стивенс-Давидовиц, 2018, с. 336]. Однако это не означает, что социальные законы будут сведены к формулам наподобие $e = mc^2$ [Стивенс-Давидовиц, 2018, с. 338].

Произошедшая цифровая революция еще не затронула в своей основе методику обучения в области социальных наук, продолжает автор. Целый массив данных, доступных для исследования, на данный момент пользуется спросом исключительно у студентов и небольшого числа экспертов. Однако это временное явление; в частности, Стивенс-Давидовиц прогнозирует радикаль-

ное расширение методологического инструментария исследований в сфере здравоохранения за счет применения сравнительного анализа «больших данных» и выявления корреляции между внешними факторами и распространением определенных заболеваний среди населения [Стивенс-Давидовиц, 2018, с. 338].

Потребовалось немало лет, чтобы прогресс в естественных науках нашел отражение в нашей повседневности (был открыт пенициллин, созданы спутники и компьютеры), продолжает автор. Подобная неспешная эволюция уготована и социальным наукам – вслед за их трансформацией благодаря анализу «больших данных». К примеру, экономист Северо-Западного университета (г. Эванстон, США) Бенджамин Ф. Джонс смог применить метод А/В тестирования для выявления наиболее эффективных учебных планов¹. Благодаря созданной им платформе EDU STAR Джонсу удалось доказать низкую эффективность игровых форм обучения для изучения математики младшими школьниками [Стивенс-Давидовиц, 2018, с. 341].

Помимо позитивной стороны анализа «больших данных», Стивенс-Давидовиц обсуждает также моральные аспекты применения этого инструментария. Большие данные позволяют прогнозировать наше поведение с «пугающей» точностью. В качестве примера автор ссылается на разработки группы американских экономистов [Netzer, Lemair, Herzenstein, 2019], посвященные проблематике кредитования. Было установлено, что частое использование потенциальным заемщиком таких слов, как «Бог», «обещаю», «оплачу», «спасибо», «больница», с высокой вероятностью гарантирует погашение кредита в срок [Стивенс-Давидовиц, 2018, с. 322].

Ссылаясь на использование подобных методов компаниями для повышения уровня продаж и государством для прогнозирования уровня преступности, Стивенс-Давидовиц задается вопросом: хотим ли мы жить в обществе, в котором по написанным нами в поисковой строке браузера словам можно будет предсказать, вернем ли мы кредит и совершим ли преступление? С его точки зре-

¹ А/В тестирование – метод исследования, при котором контрольная группа элементов сравнивается с набором тестовых групп, где изменен тот или иной элемент.

ния, это «как минимум жутковато и порой просто страшно» [Стивенс-Давидовиц, 2018, с. 323].

Очевидно, применение анализа доступной в Интернете информации, объединенной в массив «больших данных», способно повысить качество жизни современного общества и позволяет проводить максимально достоверные исследования в социальных и естественных науках. Однако при использовании подобного инструментария необходимо установить границы приемлемого с точки зрения морали. Как показал имевший место в недавнем прошлом отказ Facebook размещать политическую рекламу, последовавший вслед за слушаниями в Сенате США, которые касались участия компании *Cambridge Analytica*¹ в президентских выборах 2016 г., общество пока не готово к подобному вмешательству аналитиков «больших данных» в свою жизнь.

Список литературы

Левитт С.Д., Дабнер С.Дж. Фрикономика: Мнение экономиста-диссидента о неожиданных связях между событиями и явлениями / пер. с англ. – М.: Вильямс, 2007. – 288 с.

Стивенс-Давидовиц С. Все лгут: Поисковики, Big Data и Интернет знают о вас все / пер. с англ. Л.И. Степановой. – М.: Эксмо, 2018. – 384 с.

Netzer O., Lemaire A., Herzenstein M. When words sweat: Identifying signals for loan default in the text of loan applications / J. of marketing research. – 2019. – Vol. 56, N 6. – P. 960–980.

Silver N. The signal and the noise: Why most predictions fail – but some don't. – N.Y.: Penguin press, 2012. – 534 p.

© Павлюченко А.А., 2020

2020.02.010. ВЕР Э. ВЛИЯНИЕ ДИСТАНЦИОННОЙ РАБОТЫ НА РАБОТНИКА В ПРОФЕССИОНАЛЬНОЙ, СЕМЕЙНОЙ И ОБЩЕСТВЕННОЙ СФЕРАХ.

Реф. ст.: VAYRE É. Les incidences du télétravail sur le travailleur dans les domaines professionnel, familial et social // Le travail humain. – 2019. – Vol. 82, N 1. – P. 1–39.

¹ Cambridge Analytica (CA) – частная английская компания, которая использует технологии глубинного анализа данных для разработки стратегической коммуникации в ходе избирательных кампаний в Интернете. В 2016 г. специалисты CA сопровождали президентскую кампанию Д. Трампа.