

И.А. Щеглова

ЭТИЧЕСКИЕ И ПРАВОВЫЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ ДАННЫХ ИЗ СОЦИАЛЬНЫХ МЕДИА

Статья посвящена анализу этических и правовых аспектов исследований с использованием данных из социальных медиа для прогнозирования экономических, политических, социальных и других процессов. Цель работы – опираясь на кейсы, имеющие место в исследовательской практике, проанализировать последствия нарушений, возникающих в ходе работы с данными из социальных сетей, и предложить пути преодоления этико-правовых барьеров в использовании таких данных.

Ключевые слова: большие данные (Big data); социальные медиа; данные из соцсетей; персональные данные; этические аспекты; правовые аспекты.

В 2016 г. количество пользователей социальных медиа составило 2,8 млрд человек, что на 21% больше, чем в 2015 г. [1]. Согласно данным Brand Analytics число активных участников социальных сетей в декабре 2015 г. в России составило около 37 млн человек, ими было сгенерировано 588 млн сообщений за месяц [2]. Число активных пользователей мировой сети Facebook в сентябре 2017 г. составило 1,37 млрд человек [3]. Таким образом, интенсивное развитие информационной среды и системы глобальной коммуникации сделали Интернет, с одной стороны, массовым благом современного поколения, которое люди используют для решения своих повседневных нужд – поиска информации, обучения, общения, развлечения. С другой стороны, открытость и доступность данных о социальной жизни людей вызывает опасения, так как вся информация, которую создают и которой обмениваются пользователи внутри социальных медиа, автоматически становится достоянием общественности.

Сбор информации о респонденте без прямого обращения к нему считается нереактивным методом исследования, примером которого являются «большие данные» (Big Data) [4]. Big Data – это цифровые следы, оставляемые людьми в ходе посещения сайтов, пространственных перемещений и в результате других действий. Преимущества таких данных – объемность и оперативность [5]. В 2013 г. количество хранящейся в мире информации составило 1,2 зеттабайта, из которых всего лишь 2% занимает нецифровая информация [6].

Сегодня форумы и социальные сети открывают возможности для извлечения больших данных различного характера. Наиболее часто используемыми типами больших данных, извлекаемыми из социальных сетей, являются сообщения, публикации, репосты и лайки пользователей. Такие данные пользуются спросом среди маркетологов, экономистов, психологов, политологов, журналистов, к ним обращаются службы безопасности и представители других сфер [7–12]. Даже такой простой элемент социальных сетей, как лайк (отметка «Мне нравится»), зарекомендовал себя ценным источником информации. Например, с помощью лайков исследователи могут составить личностный портрет пользователя. Исследователи из университета Стэнфорда (США) и университета Кембриджа (Великобритания) смогли описать личностные черты пользователей на основе данных из

Facebook и доказали, что составленный ими портрет более валиден, чем личностный портрет, воспроизведенный с помощью бумажных опросников [7]. Если человек лайкает какой-то продукт или бренд в социальных сетях, маркетологи расценивают такое действие как форму выражения потребителями своих предпочтений. На основе данных из социальных сетей крупным ретейлерам удается повысить эффективность коммуникаций с потенциальными клиентами и значительно увеличить продажи [13–15]. Кроме этого, данные из социальных медиа нашли применение в медицине, например в диагностике депрессии [16, 17], оценке состояния психического здоровья [18], а также выявлении вспышек гриппа [19].

Данные из социальных сетей активно используются не только в маркетинге и медицине, но и в политике. Брексит (Brexit) и победу Дональда Трампа на выборах в США считают самыми неожиданными событиями 2016 г. Дело в том, что социологические опросы, которым привыкли доверять, дали неверные прогнозы, в то время как анализ сообщений Twitter и Facebook помог установить верные результаты американских выборов и судьбу Великобритании в Европейском Союзе [20–23]. Данные события подталкивают современную социологию общественного мнения к смене подходов к исследованиям и переходу от массовых опросов к анализу больших данных (Big Data).

Необходимо иметь в виду, что прогноз на основе больших данных может оказаться ошибочным, что впоследствии негативно скажется на имидже организации, проводившей прогностический анализ. Так, например, известно, что в 2012 г. крупная американская розничная сеть Target использовала данные об онлайн покупках клиенток, чтобы предсказать срок их беременности, а затем рассылать им предложения с товарами для беременных с целью увеличения объема продаж [9]. Однако такая «забота» вызвала вопросы у покупателей, большинство из которых не догадывались, как много известно об их частной жизни [9, 24]. В 2017 г. эксперты по анализу данных из социальных медиа неверно предсказали результаты голосования в Австралии за однополые браки, существенно занизив процент проголосовавших «за» [25, 26].

Несмотря на то, что большинство пользователей сетей не ограничивают доступ к своей демографической информации (пол, возраст, семейное положение, образование), исследователи отмечают, что при за-

полнении профиля в социальных сетях некоторые поля могут по ошибке или преднамеренно оставаться незаполненными или содержать ложную информацию о фактах биографии, интересах и предпочтениях [27–29]. Тем не менее, если пользователи скрывают данную информацию, но проявляют активность в социальных сетях, например комментируют посты, публикуют статусы, то текстовые сообщения рассматриваются как косвенные показатели их личностных характеристик, указывающих на профессию, уровень образования, политические взгляды, религиозные убеждения [5, 28]. В условиях массового сбора данных о пользователях даже такие базовые сведения, как, например, посещение сайта или совершение покупки могут служить источником более детальной информации о пользователе, чем анкетные данные [30].

Несмотря на очевидные плюсы быстро развивающейся сферы больших данных, исследователями часто игнорируется этическая составляющая данного вопроса. Согласно результатам исследования, проведенного учеными Холл и Флинн, только 36% руководителей департаментов компьютерных наук британских вузов считают, что анализ этической составляющей научных работ «очень важен», и всего лишь 1% респондентов согласился с тем, что университеты «достаточно эффективно» мониторят этическое качество исследований в сфере компьютерной инженерии [31]. Важность соблюдения этико-правовых норм в проведении исследований обостряется с появлением новых нарушений в использовании данных из социальных медиа [32–34]. В то же время растет число работ, в которых отмечается необходимость пересмотра стратегий управления информацией (создание, агрегирование, продажа данных), а также решение вопросов этики и конфиденциальности [12, 28, 35, 36].

Таким образом, с одной стороны, данные, извлекаемые из социальных сетей, дают новые возможности для развития науки, а с другой – ставят под сомнение правомерность и этичность их использования. Цель данной работы – сделать обзор этических и правовых нарушений использования данных из социальных медиа на примере кейсов, имеющих место в исследовательской практике, а также проанализировать правила, регламентирующие процедуру работы с такими данными. В работе ставятся следующие задачи:

1) описать текущее состояние этических и правовых аспектов использования данных из социальных медиа на примере российских и зарубежных работ;

2) определить спектр этических и правовых проблем, возникающих при работе с данными из социальных медиа;

3) предложить пути преодоления этико-правовых барьеров в использовании данных из социальных медиа.

Новизна работы заключается в поднятии и развитии новой и в то же время актуальной для российского научного сообщества темы – использование данных из социальных медиа в исследовательской практике для прогнозирования экономических, политических и других процессов, играющих важную роль в современном мире.

Этические аспекты. В связи со стремительным развитием и массовым распространением социальных

сетей работа с данными, извлекаемыми из них, считается новым, перспективным, но пока еще малоизученным процессом. Кроме того, этические и правовые вопросы использования данных из соцсетей не проработаны в полной мере. С одной стороны, доступность информации о поведении человека, его личностных чертах, с другой стороны, сильное желание исследователей узнать как можно больше об интересующем их объекте (например, покупателе, сотруднике) ставит серьезные вопросы к защите персональных данных и права на частную жизнь. Регистрируясь в социальных сетях, люди принимают пользовательское соглашение, тем самым дают разрешение на размещение информации о себе в открытом доступе. Однако пользователи редко задумываются о том, что может стоять за «безобидным» согласием на обработку данных. Одно дело, когда участники социальных сетей могут самостоятельно контролировать информацию, которую они делают доступной для других, другое дело, когда появляется большое количество программного обеспечения, с помощью которого можно легко получить информацию о пользователях, их активности на медиаплатформе, даже если эта информация была скрыта настройками приватности. Сегодня активно используются автоматизированные компьютерные программы – скрэперы, которые позволяют автоматически скачать и обработать большой объем информации, делая содержание сайта пригодным для использования в качестве данных [37, 38].

В 2008 г. группа исследователей из университета Гарварда в рамках проекта «Вкусы, связи и время» («Tastes, Ties, and Time») сформировала базу данных, которую можно было бы использовать для проведения идеального исследования: репрезентативная выборка, лонгитюдные данные, собранные в социальной сети Facebook и дополненные информацией от университета [39]. Исследователи ставили цель изучить взаимосвязи студентов с их социальной средой и университетской жизнью, а также произошедшие изменения за четыре года обучения. Проект был одобрен этической комиссией университета и получил финансирование от Национального научного фонда. Одним из условий получения финансирования был свободный доступ к собранным данным. Разумеется, исследователи постарались защитить участников проекта: убрали имена и ID из баз данных, обязали всех исследователей, имеющих доступ к данным, подписать соглашение о неразглашении информации. Тем не менее, несмотря на старания ученых сделать информацию конфиденциальной, университет Гарварда, зашифрованный как «неизвестный университет на северо-востоке США» («anonymous, northeastern American university»), был легко раскрыт [32]. Кроме этого, с помощью кодовой книги, находившейся в открытом доступе, можно было получить детальную информацию о выборке студентов: пол, раса, национальность, родной штат, политические взгляды, направление обучения, что позволило установить личности конкретных участников. Данное исследование получило широкий резонанс, и под нажимом общественности данные были удалены из общего доступа [40, 41].

В 2014 г. в американском научном журнале «Proceedings of the National Academy of Sciences (PNAS)» была опубликована статья «Экспериментальное доказательство массового эмоционального воздействия через социальные сети» («Experimental evidence of massive-scale emotional contagion through social networks»), в которой авторы рассказывали об эксперименте по управлению поведением и настроением людей с помощью социальных сетей [40, 41]. В своей работе исследователи доказали, что эмоциональное состояние может передаваться от одних людей другим, при этом люди не ощущают какого-то внешнего воздействия. В качестве экспериментального воздействия ученые использовали новостную ленту Facebook, манипулируя хорошими и плохими статусами и постами. Спустя некоторое время после публикации статьи в журнале редакторы высказали свои опасения по поводу этической составляющей данного исследования. Они пояснили, что статья была принята к публикации, так как работа была одобрена этической комиссией Корнельского университета и авторы статьи ссылались на Правила использования данных Facebook, согласно которым пользователи, создавая аккаунт на платформе Facebook, автоматически соглашаются, что информация о них будет доступна широкой аудитории. Однако позднее редакторы журнала указали на нарушение этических норм проведения исследования, аргументируя это тем, что участники исследования не знали, что они являются объектом изучения и, следовательно, не могли согласиться или отказаться от участия в эксперименте [40]. В то же время представитель Британского психологического общества заявили, что подобный эксперимент следует считать социально безответственным, так как своими действиями ученые нарушают автономию и достоинство людей [34]. Кроме того, в действиях научной группы увидели потенциальный вред здоровью, так как участники эксперимента подверглись воздействию негативных сообщений, которые могли вызвать отрицательные эмоции и послужить причиной нарушения психического состояния. Комментируя исследование Крамера и его команды, представитель Facebook утверждала, что «их цель заключалась в улучшении качества сервисов и стремлении сделать контент релевантным для определенных групп пользователей [Там же]. Тем не менее, как отмечает Ральф Шредер, под «улучшением качества сервисов» может подразумеваться увеличение продаж или манипуляция политическими настроениями общества, что полностью противоречит этико-правовым нормам научной практики [10. С. 4].

Таким образом, исследования, в которых собираются сетевые данные, вызывают много спорных вопросов. Основная проблема заключается в том, что они нарушают главный этический принцип – добровольное участие в исследовании. Также возникают опасения, связанные с возможным причинением вреда здоровью испытуемых и вторжением в их личную жизнь в случае деанонимизации персональных данных, собранных в ходе эксперимента. Нюрнбергский кодекс, утвержденный в 1947 г., сделал необходимым условием проведения исследования с участием чело-

века его добровольное согласие, а также информирование о возможных рисках. Принципы, представленные в Нюрнбергском кодексе, легли в основу этических норм проведения исследований в современном мире. В апреле 1979 г. в Бельмонтском докладе были закреплены этические основания для действующих законов по защите участников исследования: уважение личности, польза и справедливость [42]. Сегодня для обеспечения этических требований исследователи предлагают выполнять следующие шаги:

- предоставить пользователям подробную информацию о том, какие данные будут использоваться для исследования, и кто будет иметь к ним доступ;
- предоставить пользователям информацию о текущих результатах исследования;
- сообщить пользователям о публикации результатов исследования [28. С. 553].

Юридические аспекты. Помимо этических аспектов использования данных из социальных медиа, встает вопрос о юридических ограничениях. Широкий резонанс общественности, вызванный публикацией результатов исследования «Экспериментальное доказательство массового эмоционального воздействия через социальные сети», послужил толчком для введения юридических ограничений [33]. Центр электронной конфиденциальной информации (EPIC), американская научно-исследовательская организация, защищающая права граждан на частную жизнь, свободу выражения мнения и демократические ценности в век информации, подал судебный иск против компании Facebook за нарушение закона Федеральной торговой комиссии [43]. Основная претензия EPIC к Facebook заключалась в «скрытном и несогласованном использовании персональных данных пользователей с целью проведения над ними психологического эксперимента» [Там же]. В качестве аргумента EPIC приводит тот факт, что во время проведения эксперимента в документе об использовании данных Facebook не указал, что личная информация участников может быть использована в исследовательских целях. Таким образом, после ряда судебных разбирательств и громкой огласки руководители глобальной медиа платформы приняли требования Федеральной торговой комиссии и ввели ограничение к данным на личных страницах пользователей.

В Европе Facebook также подвергался резкой критике из-за непрозрачности политики по сбору и обработке информации о пользователях [44]. В 2015 г. исследователи междисциплинарного центра права, интеллектуальной защиты прав Левенского университета и отдела изучения медиа, информации и телекоммуникации Брюссельского свободного университета заявили, что политика конфиденциальности Facebook, обновленная в начале 2015 г., нарушает закон о защите прав потребителей Европейского союза. В ходе анализа были выделены основные замечания: 1) отсутствие валидного согласия пользователей на обработку их данных; 2) отсутствие адекватных механизмов контроля использования данных пользователей в коммерческих целях; 3) невозможность отключить функцию определения местоположения [45].

В России официально действует федеральный закон «О персональных данных» от 27.07.2006, регули-

рующий отношения, связанные с обработкой персональных данных юридическими и физическими лицами с использованием средств автоматизации, в том числе в информационно-телекоммуникационных сетях. Закон о персональных данных предусматривает [46]:

1) законное основание обработки персональных данных, в качестве которого выступает информированное, конкретное и сознательное согласие субъекта персональных данных, за исключением случаев, прямо указанных в законе;

2) ограничение обработки персональных данных только заранее установленными, законными целями (не допускается обработка персональных данных, несовместимая с целями их сбора, а равно избыточность персональных данных по отношению к целям их обработки);

3) недопустимость объединения баз данных, содержащих персональные данные, обработка которых осуществляется в целях, несовместимых между собой;

4) при обработке персональных данных должна быть обеспечена их точность и актуальность по отношению к целям обработки персональных данных;

5) хранение персональных данных должно осуществляться в форме, позволяющей определить субъекта персональных данных не дольше, чем этого требуют цели обработки персональных данных.

Однако, как показывает в своей статье А.И. Савельев, возможности, создаваемые технологиями больших данных, находятся в прямом противоречии с принципами закона, что делает право на частную жизнь и большие данные несовместимыми между собой [30]. Автор отмечает, что получение информированного согласия от респондента невозможно по следующим причинам:

а) отсутствие возможности предоставить исчерпывающий объем информации о возможных способах и целях обработки персональных данных;

б) отсутствие возможности субъекта персональных данных адекватно воспринять такую информацию;

в) отсутствие возможности взаимодействия с огромным множеством организаций, осуществляющих сбор и обработку персональных данных в современном обществе [Там же. С. 55].

Пути преодоления этико-правовых барьеров. Несмотря на то что большие данные стали одним из

перспективных направлений в научном мире, их использование открывает не только новые возможности перед учеными, но и ставит серьезные этические и юридические вопросы, на которые пока не найдены правильные ответы.

Стоит отметить, что несколько лет назад, когда ученые только начинали воплощать в жизнь свои проекты, можно было легко скачать данные из Facebook, не запрашивая разрешения пользователей на доступ к их профилю, и использовать полученную информацию для любых целей. Сегодня уже нельзя так просто получить доступ к данным на социальных платформах. Кроме того, научные журналы стали серьезнее относиться к этике исследований, проводимых на основе таких данных. Тем не менее для урегулирования этико-правовых вопросов работы с данными из социальных медиа необходимо предпринять следующие шаги:

1) переосмыслить и адаптировать к условиям медиа среды такие понятия, как персональные данные, информированное согласие, возможность выйти из исследования. Также необходимо урегулировать на правовом уровне статус персональных данных пользователей, регистрируемых в социальных сетях. Другими словами, юридически зафиксировать, являются персональные данные ресурсом, находящимся в открытом доступе, или это неотчуждаемое нематериальное благо;

2) разработать регламент предварительной экспертизы плана исследования, реализуемого в социальных медиа. Обеспечить мониторинг текущих результатов для предотвращения негативных последствий. Особое внимание необходимо уделить следующим условиям: а) полная анонимизация данных и отсутствие возможности их деанонимизировать; б) отсутствие взаимодействия между участниками выборки; в) опубликованные результаты не должны давать возможность раскрыть участников исследования;

3) разработать для обучающихся в магистратуре и аспирантуре, научных сотрудников и других лиц, вовлеченных в научные проекты, курс по этике проведения исследований (в очном или онлайн-формате), включающий раздел по использованию больших данных. Данный курс поможет исследователям правильно оценить этические аспекты будущих проектов и поддерживать эволюцию стандартов и норм в быстро меняющейся научно-технической среде.

ЛИТЕРАТУРА

1. Kemp S. Digital in 2017 Global Overview. URL: <https://hootsuite.com/de/newsroom/press-releases/digital-in-2017-report> (access date: 25.12.2017).
2. Brand Analytics. Социальные сети в России, зима 2015–2016. Цифры, тренды, прогнозы. URL: <https://blog.br-analytics.ru/socialnye-seti-v-rossii-zima-2015-2016-cifry-trendy-prognozy/> (дата обращения: 25.12.2017).
3. Key facts about Facebook. URL: <http://newsroom.fb.com/key-facts> (дата обращения: 05.01.2018).
4. Онлайн исследования в России 2.0 : сб. ст. / под ред. А.В. Шашкина, И.Ф. Девятко, С.Г. Давыдова. М. : Северо-Восток, 2010.
5. Корытникова Н.В. Online Big Data как источник аналитической информации в онлайн исследованиях // Социологические исследования. 2015. № 8. С. 14–24.
6. Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / пер. с англ. Инны Гайдюк. М. : Манн, Иванов и Фербер, 2014. 240 с.
7. Youyoua W., Kosinski M., Stillwell D. Computer-based personality judgments are more accurate than those made by humans // PNAS. 2015. № 112 (4). P. 1036–1040.
8. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior // Proceedings of the National Academy of Sciences (PNAS). 2013.
9. Duhigg C. The Power of Habit: Why we do what we do in life and business. New York : Random House, 2012.
10. Schroeder R. Big Data and the brave new world of social media research // Big Data and Society. 2014. № 1 (2). P. 1–11.

11. Lazer D., Pentland A., Adamic L.A., Aral S., Barabási A.-L., Brewer D. Van Alstyne M. Life in the network: the coming age of computational social science // *Science*. 2009. Vol. 323. P. 721–723.
12. Martin K.E. Ethical Issues in the Big Data Industry // *MIS Quarterly Executive*. 2015. № 14 (2). P. 67–85.
13. Calo M.R. Digital Market Manipulation // *The George Washington Law Review*. 2013. Vol. 82 (4). P. 995–1051.
14. Xun J., Guo B. Twitter as customer's eWOM: an empirical study on their impact on firm financial performance // *Internet Research*. 2017. Vol. 27 (5). P. 1014–1038.
15. Brynjolfsson E., Hitt L., Kim H. Strength in Numbers: How does data-driven decision-making affect firm performance? // *ICIS 2011 Proceedings*. 2011. URL: <https://aisel.aisnet.org/icis2011/proceedings/economicvalueIS/13> (access date: 20.12.2017).
16. Moreno M.A., Jelenchick L.A., Egan K.G., Cox E., Young H., Gannon K.E., Becker T. Feeling bad on Facebook: depression disclosures by college students on a social networking site // *Depression and Anxiety*. 2011. Vol. 28. P. 447–455.
17. Reece A.G., Danforth C.M. Instagram photos reveal predictive markers of depression // 6:15. 2016.
18. Conway M., O'Connor D. Social media, big data, and mental health: current advances and ethical implications // *Current Opinion in Psychology*. 2016. Vol. 9. P. 77–82.
19. Li J., Cardie C. Early Stage Influenza Detection from Twitter // *arXiv preprint arXiv:1309.7340*. 2013.
20. Шалак В. Potus 2016. URL: <http://vaal.ru/show.php?id=281> (дата обращения: 05.01.2018).
21. Perez S. Analysis of social media did a better job at predicting Trump's win than the polls // *Techcrunch*. 2016. URL: <https://techcrunch.com/2016/11/10/social-media-did-a-better-job-at-predicting-trumps-win-than-the-polls/> (access date: 05.01.2018).
22. Ceron A., Curini L., Iacus S. How pollsters could use social media data to improve election forecasts // *The Washington Post*. 2016. URL: https://www.washingtonpost.com/news/monkey-cage/wp/2016/12/21/using-these-4-steps-pollsters-could-use-social-media-data-to-improve-election-forecasts/?utm_term=.a9d5a334b373 (access date: 05.01.2018).
23. Celli F., Stepanov E., Poesio M., Riccardi G. Predicting Brexit: Classifying Agreement is Better than Sentiment and Pollsters // *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*. Osaka, Japan, December, 2016. P. 110–118.
24. Lubin G. The Incredible Story of How Target Exposed A Teen Girl's Pregnancy. 2012. URL: <http://www.businessinsider.com/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2> (access date: 04.01.2018).
25. Tuffley D., Stantic B. Social media study points to a close result in the same-sex marriage vote // *The conversation*. 2017. URL: <https://theconversation.com/social-media-study-points-to-a-close-result-in-the-same-sex-marriage-vote-84436> (access date: 05.01.2018).
26. Han M. Social media poll predicts narrow "No" win in same-sex marriage survey // *Financial Review*. 2017. URL: <http://www.afr.com/news/politics/social-media-poll-predicts-narrow-no-win-in-samesex-marriage-survey-20171101-gzd6rv> (access date: 04.01.2018).
27. Коршунов А., Белобородов И., Бузун Н., Аванесов В., Пастухов Р., Чихрадзе К., Козлов И., Гомзин А., Андрианов И., Сысоев А., Ипатов С., Филоненко И., Чуприна К., Турдаков Д., Кузнецов С. Анализ социальных сетей: методы и приложения // *Труды Института системного программирования РАН*. 2014. Т. 26, вып. 1. С. 439–456.
28. Kosinski M., Matz S., Gosling S., Popov V., Stillwell D. Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines // *American Psychologist*. 2015.
29. Yang Z., Wilson C., Wang X., Gao T., Zhao, B. Y., Dai Y. Uncovering social network Sybils in the wild // *Proceedings of the ACM SIGCOMM conference on Internet measurement*. New York, NY : ACM Press, 2011.
30. Савельев А.И. Проблемы применения законодательства о персональных данных в эпоху «Больших данных» (Big Data) // *Право. Журнал Высшей школы экономики*. 2015. № 1. С. 43–66.
31. Hall T., Flynn V. Ethical issues in software engineering research: a survey of current practice // *Empirical Software Engineering*. 2001. Vol. 6. P. 305–317.
32. Zimmer M. But the data is already public: on the ethics of research in Facebook // *Ethics and Information Technology*. 2010. № 12. P. 313–325.
33. Kramera A.D., Guillory J.E., Hancock J.T. Experimental evidence of massive-scale emotional contagion through social networks // *PNAS*. 2014. № 111 (24). P. 8788–8790.
34. Facebook's 'experiment' was socially irresponsible // *Guardian*. 2014. URL: <https://www.theguardian.com/technology/2014/jul/01/facebook-socially-irresponsible> (access date: 17.07.2017).
35. Norval C., Henderson T. Contextual Consent: Ethical Mining of Social Media for Health Research // *Proceedings of the WSDM 2017, Workshop on Mining Online Health Reports*. Cambridge, 2017.
36. Israel M., Hay I. *Research Ethics for Social Scientists: Between Ethical Conduct and Regulatory Compliance*. London : Sage, 2006.
37. Barbier G., Liu H. Data Mining in Social Media // Aggarwal C. (eds) *Social Network Data Analytics*. Boston, MA : Springer, 2011.
38. Hogan B. Analyzing Social Networks via the Internet // *The SAGE Handbook of Online Research Methods* Nigel Fielding / R.M. Lee & Grant Blank (eds). 2008.
39. Lewis K., Kaufman J., Gonzalez M., Wimmer A., Christakis N. Tastes, Ties, and time: A new social network dataset using Facebook.Com // *Social Networks*. 2008. No. 30 (4). P. 330–342.
40. Parry M. Harvard Researchers Accused of Breaching Students' Privacy // *The Chronicle of Higher Education*. 2011. URL: <https://www.chronicle.com/article/Harvards-Privacy-Meltdown/128166> (access date: 05.01.2018).
41. Kaiser T. Harvard Sociologists in Hot Water Over Facebook Study; Student Privacy Called into Question // *Dailytech*. 2011. URL: <http://www.dailytech.com/Harvard+Sociologists+in+Hot+Water+Over+Facebook+Study+Student+Privacy+Called+into+Question/article22116.htm> (access date: 04.01.2018).
42. The Belmont Report. URL: <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html> (access date: 10.07.2017).
43. In re: Facebook (Psychological Study). EPIC, 2014. URL: <https://epic.org/privacy/internet/ftc/facebook/psycho/> (access date: 17.07.2017).
44. Lomas N. Facebook's Data Protection Practices Under Fresh Fire In Europe // *TechCrunch*. 2015. URL: <https://techcrunch.com/2015/02/23/facebook-ad-network/> (access date: 2.01.2018).
45. Van Alsenoy B., Verdoodt V., Heyman R., Ausloos J., Wauters E. From social media service to advertising network: a critical analysis of Facebook's Revised Policies and Terms. 2015. URL: <https://www.law.kuleuven.be/citip/en/news/item/facebook-revised-policies-and-terms-v1-1.pdf> (access date: 20.12.2017).
46. О персональных данных : Федеральный закон РФ. URL: http://www.consultant.ru/document/cons_doc_LAW_61801/ (дата обращения: 13.07.2017).

Статья представлена научной редакцией «Социология и политология» 18 января 2018 г.

ETHICAL AND LEGAL ASPECTS OF SOCIAL MEDIA DATA USAGE

Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal, 2018, 431, 81–87.

DOI: 10.17223/15617793/431/10

Irina A. Shcheglova, Higher School of Economics (Moscow, Russian Federation). E-mail: ishcheglova@hse.ru; shcheglova.i@mail.ru

Keywords: Big Data; social media; social networking data; personal data; ethical aspects; legal aspects.

This paper discusses the ethical and legal usage of social networking data (for instance, Facebook, VKontakte, Twitter, YouTube and other websites) enabled to answer vital research questions, generating economic, political and social forecasts, as well as understanding other processes. It aims to analyze the ethical and legal consequences of data misuse which took place in research practice and offers steps to prevent ethical and legal barriers that might arise while working with social networking data. The paper poses the following research tasks: 1) to describe the current state of ethical and legal aspects of social networking data usage taking into account existing Russian and international research; 2) to identify the circle of ethical and legal concerns that might emerge while working with social networking data; 3) to provide recommendations aimed at preventing ethical and legal misuse of such data. As it is shown in the paper, the main issue with the usage of such data is that it undermines the main ethical principle – voluntary participation in research. In addition, the tempting openness of social networking data can encourage the infringement of privacy rights, threaten personal data protection and harm the mental health of users. In spite of all efforts countries put to handle legal questions related to the usage of Big Data, current practice shows that possibilities developed by Big Data undermine principles of law and make people's right to privacy and such data incompatible. In the paper, the author suggests the following steps aimed at preventing ethical and legal violations of social media data usage: 1) to reconsider and adapt to social media environment such concepts as personal data, informed consent, ability of people to opt in and out of an experiment; 2) to develop detailed regulations of preliminary ethical and legal expertise of research proposals with social networking data and monitor results of ongoing research to prevent potential harm which might be caused by the usage of such data; 3) to develop a course (face-to-face or online) on research ethics with modules on Big Data usage for students, researchers and people who can be potentially involved in any research field.

REFERENCES

1. Kemp, S. (2017) *Digital in 2017 Global Overview*. [Online] Available from: <https://hootsuite.com/de/newsroom/press-releases/digital-in-2017-report>. (Accessed: 25.12.2017).
2. Brand Analytics. (2016) *Sotsial'nye seti v Rossii, zima 2015–2016. Tsifry, trendy, prognozy* [Social networks in Russia, winter 2015–2016. Figures, trends, forecasts]. [Online] Available from: <https://blog.br-analytics.ru/socialnye-seti-v-rossii-zima-2015-2016-tsifry-trendy-prognozy/>. (Accessed: 25.12.2017).
3. Facebook. (n.d.) *Key facts about Facebook*. [Online] Available from: <http://newsroom.fb.com/key-facts>. (Accessed: 05.01.2018).
4. Shashkin, A.V., Devyatko, I.F. & Davydova, S.G. (eds) (2010) *Onlayn issledovaniya v Rossii 2.0* [Online research in Russia 2.0]. Moscow: Severo-Vostok.
5. Korytnikova, N.V. (2015) Online Big Data kak istochnik analiticheskoy informatsii v onlayn issledovaniyakh [Online Big Data as a source of analytical information in online research]. *Sotsiologicheskie issledovaniya – Sociological Studies*. 8. pp. 14–24.
6. Mayer-Schonberger, V. & Cukier, K. (2014) *Bol'shie dannye. Revolyutsiya, kotoraya izmenit to, kak my zhivem, rabotaem i myslim* [Big Data. A Revolution That Will Transform How We Live, Work, and Think]. Translated from English by Inna Gaydyuk. Moscow: Mann, Ivanov i Ferber.
7. Youyoua, W., Kosinski, M. & Stillwell, D. (2015) Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences (PNAS)*. 112 (4). pp. 1036–1040. DOI: 10.1073/pnas.1418680112
8. Kosinski, M., Stillwell, D. & Graepel, T. (2013) Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences (PNAS)*. 110 (15). DOI: 10.1073/pnas.1218772110
9. Duhigg, C. (2012) *The Power of Habit: Why we do what we do in life and business*. New York: Random House.
10. Schroeder, R. (2014) Big Data and the brave new world of social media research. *Big Data and Society*. 1 (2). pp. 1–11. DOI: 10.1177/2053951714563194
11. Lazer, D. et al. (2009) Life in the network: the coming age of computational social science. *Science*. 323. pp. 721–723. DOI: 10.1126/science.1167742
12. Martin, K.E. (2015) Ethical Issues in the Big Data Industry. *MIS Quarterly Executive*. 14 (2). pp. 67–85.
13. Calo, M.R. (2013) Digital Market Manipulation. *The George Washington Law Review*. 82 (4). pp. 995–1051.
14. Xun, J. & Guo, B. (2017) Twitter as customer's eWOM: an empirical study on their impact on firm financial performance. *Internet Research*. 27 (5). pp. 1014–1038.
15. Brynjolfsson, E., Hitt, L. & Kim, H. (2011) Strength in Numbers: How does data-driven decision-making affect firm performance? *ICIS Proceedings*. [Online] Available from: <https://aisel.aisnet.org/icis2011/proceedings/economicvalue/13>. (Accessed: 20.12.2017).
16. Moreno, M.A. et al. (2011) Feeling bad on Facebook: depression disclosures by college students on a social networking site. *Depression and Anxiety*. 28. pp. 447–455. DOI: 10.1002/da.20805
17. Reece, A.G. & Danforth, C.M. (2016) Instagram photos reveal predictive markers of depression. *EPJ Data Science*. 6:15. DOI: 10.1140/epjds/s13688-017-0110-z
18. Conway, M. & O'Connor, S. (2016) Social media, big data, and mental health: current advances and ethical implications. *Current Opinion in Psychology*. 9. pp. 77–82. DOI: 10.1016/j.copsyc.2016.01.004
19. Li, J. & Cardie, C. (2013) *Early Stage Influenza Detection from Twitter*. arXiv preprint arXiv:1309.7340.
20. Shalak, V. (2016) *Potus*. [Online] Available from: <http://vaal.ru/show.php?id=281>. (Accessed: 05.01.2018). (In Russian).
21. Perez, S. (2016) Analysis of social media did a better job at predicting Trump's win than the polls. *TechCrunch*. [Online] Available from: <https://techcrunch.com/2016/11/10/social-media-did-a-better-job-at-predicting-trumps-win-than-the-polls/>. (Accessed: 05.01.2018).
22. Ceron, A., Curini, L. & Iacus, S. (2016) How pollsters could use social media data to improve election forecasts. *The Washington Post*. [Online] Available from: https://www.washingtonpost.com/news/monkey-cage/wp/2016/12/21/using-these-4-steps-pollsters-could-use-social-media-data-to-improve-election-forecasts/?utm_term=.a9d5a334b373. (Accessed: 05.01.2018).
23. Celli, F., Stepanov, E., Poesio, M. & Riccardi, G. (2016) Predicting Brexit: Classifying Agreement is Better than Sentiment and Pollsters. *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*. Osaka, Japan: PEOPLES@COLING. pp. 110–118.
24. Lubin, G. (2012) *The Incredible Story of How Target Exposed A Teen Girl's Pregnancy*. [Online] Available from: <http://www.businessinsider.com/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2>. (Accessed: 04.01.2018).
25. Tuffley, D. & Stantic, B. (2017) Social media study points to a close result in the same-sex marriage vote. *The Conversation*. [Online] Available from: <https://theconversation.com/social-media-study-points-to-a-close-result-in-the-same-sex-marriage-vote-84436>. (Accessed: 05.01.2018).
26. Han, M. (2017) Social media poll predicts narrow "No" win in same-sex marriage survey. *Financial Review*. [Online] Available from: <http://www.afr.com/news/politics/social-media-poll-predicts-narrow-no-win-in-samesex-marriage-survey-20171101-gzd6rv>. (Accessed: 04.01.2018).
27. Korshunov, A. et al. (2014) Analiz sotsial'nykh setey: metody i prilozheniya [Social network analysis: methods and applications]. *Trudy Instituta sistemnogo programmirovaniya RAN – Proceedings of the Institute for System Programming*. 26:1. pp. 439–456.
28. Kosinski, M., Matz, S., Gosling, S., Popov, V. & Stillwell, S. (2015) Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines. *American Psychologist*. 70(6). pp. 543–556. DOI: 10.1037/a0039210
29. Yang, Z. et al. (2011) Uncovering social network Sybils in the wild. *Proceedings of the ACM SIGCOMM conference on Internet measurement conference*. New York, NY: ACM Press. DOI: 10.1145/2556609

30. Savel'ev, A.I. (2015) The Issues of Implementing Legislation on Personal Data in the Era of Big Data. *Pravo. Zhurnal Vysshey shkoly ekonomiki – Law. Journal of the Higher School of Economics*. 1. pp. 43–66. (In Russian).
31. Hall, T. & Flynn, V. (2001) Ethical issues in software engineering research: a survey of current practice. *Empirical Software Engineering*. 6. pp. 305–317. DOI: 10.1023/A:1011922615502
32. Zimmer, M. (2010) “But the data is already public”: on the ethics of research in Facebook. *Ethics and Information Technology*. 12. pp. 313–325. DOI: 10.1007/s10676-010-9227-5
33. Kramera, A.D., Guillory, J.E. & Hancock, J.T. (2014) Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*. 111 (24). pp. 8788–8790. DOI: 10.1073/pnas.1320040111
34. The Guardian. (2014) Facebook's ‘experiment’ was socially irresponsible. *The Guardian*. [Online] Available from: <https://www.theguardian.com/technology/2014/jul/01/facebook-socially-irresponsible>. (Accessed: 17.07.2017).
35. Norval, C. & Henderson, T. (2017) Contextual Consent: Ethical Mining of Social Media for Health Research. *Proceedings of the WSDM 2017, Workshop on Mining Online Health Reports*. Cambridge.
36. Israel, M. & Hay, I. (2006) *Research Ethics for Social Scientists: Between Ethical Conduct and Regulatory Compliance*. London: Sage.
37. Barbier, G. & Liu, H. (2011) Data Mining in Social Media. In: Aggarwal, C. (eds) *Social Network Data Analytics*. Boston, MA: Springer.
38. Hogan, B. (2008) Analyzing Social Networks via the Internet. In: Fielding, N., Lee, R.M. & Blank, G. (eds) *The SAGE Handbook of Online Research Methods*. London; Beverly Hills, CA: Sage.
39. Lewis, K. et al. (2008) Tastes, Ties, and time: A new social network dataset using Facebook.com. *Social Networks*. 30 (4). pp. 330–342. DOI: 10.1016/j.socnet.2008.07.002
40. Parry, M. (2011) Harvard Researchers Accused of Breaching Students' Privacy. *The Chronicle of Higher Education*. [Online] Available from: <https://www.chronicle.com/article/Harvards-Privacy-Meltdown/128166>. (Accessed: 05.01.2018).
41. Kaiser, T. (2011) Harvard Sociologists in Hot Water Over Facebook Study; Student Privacy Called into Question. *Dailytech*. [Online] Available from: <http://www.dailytech.com/Harvard+Sociologists+in+Hot+Water+Over+Facebook+Study+Student+Privacy+Called+into+Question/article22116.htm>. (Accessed: 04.01.2018).
42. HHS.gov. (1979) *The Belmont Report*. [Online] Available from: <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>. (Accessed: 10.07.2017).
43. EPIC. (2014) *In re: Facebook (Psychological Study)*. [Online] Available from: <https://epic.org/privacy/internet/ftc/facebook/psycho/>. (Accessed: 17.07.2017).
44. Lomas, N. (2015) Facebook's Data Protection Practices Under Fresh Fire In Europe. *TechCrunch*. [Online] Available from: <https://techcrunch.com/2015/02/23/facebook-ad-network/>. (Accessed: 2.01.2018).
45. Van Alsenoy, B. et al. (2015) *From social media service to advertising network: a critical analysis of Facebook's Revised Policies and Terms*. [Online] Available from: <https://www.law.kuleuven.be/citip/en/news/item/facebook-revised-policies-and-terms-v1-1.pdf>. (Accessed: 20.12.2017).
46. Consultant.ru. (2006) *O personal'nykh dannykh: Federal'nyy zakon RF* [On the personal data: the Federal law of the Russian Federation]. [Online] Available from: http://www.consultant.ru/document/cons_doc_LAW_61801/. (Accessed: 13.07.2017).

Received: 18 January 2018