

УДК 1/16. 001.167.1.004.174

БОЛЬШИЕ ДАННЫЕ И ЭТИКА: ТЕРРИТОРИЯ ОБСУЖДЕНИЯ

Багиров М.Б., Михайлова Т.Л.

ФГБОУ ВО «Нижегородский государственный технический университет им. Р.Е. Алексеева», Нижний Новгород, Россия (603950, Нижний Новгород, ГСП-41, ул. Минина, 24), e-mail: tmichailova2012@yandex.ru

Статья посвящена исследованию стремительно растущего феномена больших данных. Показано его влияние на различные сферы общества – экономику, политику, безопасность, образование, науку, здравоохранение. Выявлена ключевая проблема использования метода больших данных, суть которой состоит в несоответствии норм рецензирования этичности исследования. Рассмотрено, что с применением методов больших данных понятийные и инфраструктурные ресурсы, выработанные ранее, перестали быть эффективными для рецензирования, пропуская широкий спектр рисков. Обозначена причина проблемы: информатика и прикладная математика, будучи ответственными за прогресс больших данных, дистанцируются от этики исследований. Итогом статьи является вывод о необходимости пересмотра общих, сугубо технических подходов к большим данным. Приведены рекомендации преодоления понятийного и институционального разрыва между большими данными и современными нормами исследовательской этики.

Ключевые слова: большие данные, эпистемиология, информатика, этика, этические правила, обработка информации, наука о данных, персонализация данных, обезличивание данных, гибридные пространства.

BIG DATA AND ETHICS: TERRITORY FOR DISCUSSION

Bagirov M.B., Mikhailova T. L.

Nizhny Novgorod State Technical University n.a. R.E. Alekseev, Nizhny Novgorod, Russia (603950, Nizhny Novgorod, GSP-41, street Minina, 24), e-mail: tmichailova2012@yandex.ru

The article deals with the study of a rapidly growing phenomenon of big data. Its influence on various spheres of society such as economy, politics, security, education, science, and health care is analysed. The key problem of using the big data method is revealed. It is the problem of maladjustment of norms for reviewing the research ethics. It is considered that with the use of big data methods the conceptual and infrastructural resources developed earlier have ceased to be effective for reviewing, as they entail a wide range of risks. The cause of the problem is identified - computer science and applied mathematics, which are mainly responsible for the progress in the field of big data, have distanced themselves from the ethics of research. The result of the article is the conclusion about the need to revise the basic, purely technical approaches to big data. The recommendations for overcoming the conceptual and institutional gap between big data and modern norms of research ethics are proposed.

Key words: big data, epistemology, informatics, ethics, ethical rules, information processing, data science, personalization of data, depersonalization of data, hybrid spaces..

Актуальность темы обусловлена потребностью переосмысления стремительно набирающего обороты феномена больших данных. Действительно, некий ажиотаж, образовавшийся вокруг феномена больших данных, породив «турбулентности» как вокруг интерпретации больших данных, так и факта неожиданно обнаружившегося приближения дисциплин – информатики и прикладной математики, связанных с прогрессом в методах больших данных, – к гуманитарной сфере. До настоящего времени этот феномен рассматривался как чисто технический, безотносительно гуманитарных последствий, но в последние лет пять, даже

магистры и аспиранты технических специальностей, все чаще обращают пристальное внимание на гуманитарную составляющую больших данных [1; 3]. Однако возможности, предоставляемые большими данными, выходят за рамки просто технического, унитарного феномена. Компания Google сумела захватить рынок рекламы в интернете, не обладая продвинутой знаниями в области маркетинга, ими просто были применены математические методы [4].

Неожиданное сближение технического и гуманитарного инициировало обсуждение ряда аспектов. *Первый – онтологический аспект, связанный с архитектурой феномена сложности.* Быстрорастущие в последнее десятилетие большие данные, их *вездесущность* – проникновение во все, без исключения сферы жизни общества: экономика, правопорядок, политика, образование, наука, безопасность, политика. Сферы, в которых они «живут» своей жизнью, порождая проблему их *использования*.

Второй – эпистемологический аспект, тесно связанный с прагматическим аспектом. Феномен осмысления этой растущей эволюционирующей сложности, конфигурации и сплетения которой не всегда «схватываются» исследователем. Непрерывно бесконечно собираемые большие данные, организующиеся в сетях, обладающих высокой степенью гибкости, порождают проблемы перед исследователями, связанные с уже не техническими сложностями, а *сложностями рефлексивно-методологического характера*. Поясним это. Способность анализировать наборы данных из очень разрозненных контекстов и генерировать новые непредвиденные знания создает *основу, как для власти, так и для рисков исследований больших данных* и более широкого научного анализа данных. Перед нами переплетение эпистемологических и праксологических проблем, невозможность дифференцированного анализа этого явления – некий «спрут», неразъединимый даже для теоретического осмысления. Действительно, на наших глазах, очень быстро, инфраструктура и методы больших данных привели к значительным интеллектуальным и организационным изменениям для многих академических дисциплин, государственных органов, благотворительных, некоммерческих организаций, частных предприятий. Учитывая их недавний масштабный охват, большие данные по-прежнему влияют на последствия того, как заинтересованные стороны, имея доступ и обладая властью, распределяют социальные и политические блага. Новые промышленные, образовательные, и правительственные инвестиции в науку о данных и искусственный интеллект подчеркивают то, как большие данные сворачивают исторически сегментированные технические подходы. Из-за шумихи вокруг новой конструируемой реальности этих событий-переплетений НЕ сразу прояснилось становящееся очевидным сегодня то, что большие данные *поднимают этические вопросы, требующие обсуждения*.

*Итак, третий аспект – аксиологический, связанный с феноменом элиминирования этики, ее недооценки. Этическая составляющая больших данных, требующая подключения экспертного сообщества, только начинает привлекать научное сообщество. Проясним этот момент. Большие данные обходят многие неофициальные формы регулирования этики, существующие в других сообществах в области науки и техники. Например, три дисциплины, информирующие зарождающуюся область науки о данных (*информатика, физика и прикладная математика*), уже давно рассматриваются за пределами проблем, связанных с человеческими субъектами, ибо их вклад исторически были связаны с системами, а не людьми. Эти дисциплины нейтральны: они *вне круга забот об этике*. В результате, содержание наборов данных в игре считается не относящимся к основным вопросам этики. Однако, поскольку большие данные неизбежно привлекают эти дисциплины ближе к личным уязвимым данным человека, в области науки о данных обнаруживается, что у нее нет учебной программы по этике или учебных материалов, разработанных для решения этических задач. Наблюдается отставание, которое может перерасти в настоящую пропасть. Попытки саморегулирования в этой области были в значительной степени ситуативными, такими как *экспертные группы по этике*. Так как этих областях не требовалось практическое преодоление этических требований, им часто не хватает доступа к педагогическим ресурсам по исследовательской этике, широко распространенным в других областях. *Концептуальные, нормативные и институциональные ресурсы исследовательской этики*, разработанные в течение последних 70 лет, были основаны на предположениях о практике исследований данных о людях, которые иногда нелегко применять к аналитике данных.*

*Итак, проблема, которая инициирует наше исследование и связана с перечисленными выше тремя философско-методологическими аспектами, могла бы быть сформирована как проблема несоответствия технико-инструментального использования больших данных и этической составляющей применения. Если взять более широко эту проблему, то ее можно обозначить как *понятийный и институциональный разрыв* между большими данными и современными нормами исследовательской этики. Эти три аспекта, выделенные нами, не снимают и проблему отсутствия однозначного определения.*

Приведем наиболее типичные рассуждения. Многие ученые отмечают, что необходимо рассматривать этот феномен в широком *контексте понятийных изменений*. К примеру, Кейт Кройуфорд, Кейт Милтнер, и Мэри Грей (2014) считают, что нельзя позволять технологическим средствам совершать слишком много работы по трактовке при анализе больших данных, так как это приводит к узкому *взгляду на большие данные как унитарный, чисто технический феномен*. По их мнению, вместо этого мы должны рассматривать

большие данные как «*мифологический артефакт*», и стремиться проследить способы, которыми большие данные воспроизводят и изменяют знания и способы управления [8]. Джеффри Брукер утверждает, что «чистые данные» – это оксюморон, поскольку все базы данных сильно встроены в контекст, и имеют *свои временные, пространственные и материальные особенности* которые требуют более тщательного рассмотрения. [5] Подобный подход идет вразрез с текущим представлением многих ученых и деятелей индустрии о том, что большие данные это готовый для масштабного использования продукт. Некоторые ученые даже успели объявить большие данные «концом теории» и началом науки свободной от гипотез [4]. Как видим, последнее проблематизирует саму науку. Не идет ли речь о реформатировании науки? Или о становлении нового этапа науки? Возможно эти философско-методологические рефлексии выводят на новые грани научной рациональности. «Большие данные» – это не просто характеристика объема данных. Выросли и наши технические возможности по их обработке и передаче. С помощью алгоритмического анализа возможно связывать разбросанные по времени и контексту данные, устанавливая неожиданные связи. Данах Бойд и Кейт Кроуфорд (2012) утверждают, что подобное может привести к апофении, «нахождению закономерностей там, где их нет, просто по причине того, что колоссальные объемы данных могут «излучать» связи по всем направлениям» [6]. Инфраструктуры больших данных создавались с целью иметь возможность сравнить любые данные с любыми данными, что неизбежно приводит к нахождению корреляций, которые не имеют никакого смысла, однако успевают оказать влияние на все сферы, от маркетинга до законодательства. Отмеченный выше онтологический аспект исследования проявляет себе в как раз в такого рода корреляциям.

Помимо этого, большие данные расширяют наше понимание этичности исследования, сдвигая фокус от категорий прямого вреда, в сторону косвенных, таких как длительная слежка, индивидуальная и групповая дискриминация, а также «*прогностические нарушения конфиденциальности*» [8]. Очень широкий резонанс вызвало «исследование Facebook об эмоциональном заражении» [10]. Оно показало, что, изменяя эмоциональный настрой постов в новостной ленте можно влиять на эмоциональный настрой постов самих пользователей. Подобное исследование вызвало множество споров о том, насколько подобное вмешательство этично, равно как и о том, насколько при этом подобные исследования важны [11]. Стало очевидно, что при оценке этичности необходимо тщательнее смотреть на то, *как и зачем* используются данные в исследовании, и с какими наборами данных совмещаются.

Целесообразно в свете рассмотренного выше приведем некоторые рекомендации. Первое. Для борьбы с понятийным и институциональным разрывом между большими данными и современными нормами исследовательской этики, необходимо выработать ряд

рекомендаций и путей популяризации обсуждений данного вопроса. Приведем эти *рекомендации* согласно сферам, в которых большие данные «живут».

Вопросы политики

1. Правила рецензирования должны напрямую адресовать регулирование науки о данных. Очевидно, что подход «прошедший публикацию, значит безопасный» – доказал свою неэффективность. Большие данные по своей природе способны производить *принципиально новые знания* из уже существующих массивов, создавая *новые категории нанесения вреда конфиденциальности субъектов*.

2. Поддерживать появление *новых подходов к рецензированию этичности* в академических и производственных кругах. Необходимо пробовать новые подходы, способные привлечь во внимание *потенциал нанесения группового вреда*, возникающего при исследовании больших данных, учесть разницу во власти между исследователями и субъектами и где, возможно, включить отзывы от задействованного населения. Например, комиссии могут быть более приспособлены для условий и норм науки о данных, если в состав комиссий включать *ad hoc* представителей с экспертным мнением в исследуемой области.

3. Разработать механизмы оценки этичности исследований с использованием больших данных. Необходимо больше вложений в разработку средств технического вмешательства в возникающие на практике барьеры для проверки этики, средств подобных алгоритмической проверке и формальной верификации величин. Лучшее понимание организационных, культурных и технических решений в разработке систем больших данных так же важны и требуют дальнейшего изучения. Важно рассмотреть критические точки столкновения исследования и практики больших данных, там, где рассмотрение этических проблем будет наиболее плодотворным. В индустриях это может быть QA-тестирование. В академических кругах - разработка протоколов для рецензентов и встреч комитетов.

Педагогические меры

1. Создать и распространить высококачественные *тематические исследования этики данных*. Тематические исследования – это ценный педагогический ресурс в *прикладной этике*, однако такими исследованиями мало кто занимается. Стоит в первую очередь сделать акцент на исследованиях, отражающих реальные дилеммы, стоящие перед исследователями. Исследования в других дисциплинах показали, что тематическим исследованиям свойственно не учитывать сложность решений, принимаемых инженерами, по мере того, как они адресуют этические, общественные и политические динамики своих исследований и практик.

2. *Интегрировать подходы к обучению этики в учебные программы науки о данных.*

Поскольку прогресс исследований больших данных потребует разнообразных подходов, рекомендуется поддерживать преодоление интеллектуальных и институциональных разрывов между дисциплинами. Независимо от того, идет ли рассмотрение этики отдельным модулем либо ложатся в основу курса, совершенно ясно, что *этика должна стать краеугольным камнем обучения большим данным*.

3. Способствовать укоренению деятельности, направленной на этику в профессиональных ассоциациях. Чтобы обеспечить соблюдение этики в исследованиях возможно воспользоваться *практикой интеграции этики в процесс обучения*, аккредитации и практики, которая так распространена в других сферах, к примеру, в медицине. Несмотря на то, что такие усилия обычно направлены на людей, занимающихся в прикладном направлении, влияние такого подхода заметно и в исследовательском сообществе.

Развитие культуры взаимодействия

1. Создать *гибридные пространства для взаимодействия с этикой*. Специалисты по этике данных редко пересекаются с инженерами в этой области. Поэтому необходимо *поощрять сотрудничество* и неформальные связи в этих областях.

2. Разработать модели внешней и внутренней регуляции этики в индустрии. В крупных индустриях проблема этики встает довольно остро, по причине того, что основной целью является не производство знания, а создание наиболее прибыльного продукта. *Ryan Calo* предложил *концепцию «пользовательского комитета рецензирования»*, способного позволить пользователям крупных сервисов оказывать влияние на принятие решения о том, как могут быть использованы их данные [7].

3. Задать стандарты для ответственного обмена информацией между секторами деятельности в области больших данных. Уже существуют организации, занимающиеся продвижением этичности анализа данных. Подобный подход нужно направить и на области довольно деликатных корпоративных данных. Вопрос сложный, так как, с одной стороны, отсутствие контроля может привести к серьезным последствиям, к примеру, нарушению конфиденциальности. Однако наличие слишком строгого контроля может привести к отсечению широкой общественности и большинства ученых от исследований и их результатов, к элитизму со стороны исследователей, обладающих более широким доступом к данным, что не может не сказаться на качестве и этичности проводимых исследований, да и на обстановке в академической сфере вообще.

Области для дальнейшего исследования.

Следующие вопросы необходимо рассмотреть более подробно, так как ответы на них до сих пор не ясны.

1. *Должна ли сфера исследования человеческих данных считаться исследованием с*

участием субъектов-людей? Несоответствие стандартов этики текущему положению дел в исследованиях больших данных очевидно, однако, что с этим делать до сих пор неясно. В сердце этого разрыва лежит вопрос, нужно ли считать исследования с участием человеческих данных исследованиями с участием субъектов-людей, со всеми вытекающими последствиями. Вопрос *как глубоко философский, так и завязанный на практике.*

2. *Какие риски способны породить исследования в области данных?* Нам очень мало известно о *конкретных рисках*, которые несет за собой применение в исследованиях методик больших данных. Исследователи конфиденциальности и безопасности данных показали, *как* подобные риски возникают, однако информации о том, как подобные риски измерять и как предотвратить их развитие все еще очень мало.

3. Как мы можем подготовиться к появлению рисков, связанных с открытостью данных, если мы даже не знаем вместе с какими массивами данных нам предстоит обрабатывать их в будущем? Данный вопрос обозначает очень важную область для дальнейших исследований, необходим надежный способ определять и измерять риски, несмотря на такую неопределенность [12].

Подведем итог. Нами были рассмотрены вопросы необходимости широких изменений *в сфере норм исследовательской этики и появления институционального ресурса, учитывающего опыт и методы широкомасштабного анализа данных.* Ключевые на текущий момент представления были поставлены под вопрос с появлением больших данных, и наши *инструменты контроля этичности* в исследованиях требуют серьезной работы по настройке. Однако, не стоит уходить и в радикальное перекраивание существующих стандартов контроля. Появление широкомасштабной аналитики предоставляет серьезную пищу для переосмысления и уточнения *базовых представлений об этике, справедливости и ответственности*, вместе с возможностью направить ход научно-технической революции в правильное русло. Ученые, спонсоры исследований и рецензенты должны найти путь к тому, чтобы наука о больших данных учитывала ключевые проблемы этики. Территория диалога, где главным героем являются нормы исследовательской этики, – есть то коммуникативное пространство, в котором коммуникационное и коммуникативное его измерения [2] представлены в действительной своей слитности.

Список литературы

- [1]. Генералова, Е.С. Понятие «Большие данные» в контексте коммуникативистики / Е.С. Генералова, Т.Л. Михайлова // Будущее технической науки: сборник материалов XIII Международной научно-технической конференции. 2014. С. 529-530.

- [2]. Михайлова, Т.Л. Проблематизация теоретических основ коммуникативистики // Вестник НГТУ им. Р.Е. Алексеева. Серия: Управление в социальных системах. Коммуникативные технологии. 2008. № 1. С. 26-38.
- [3]. Палицына, С.С. Роль больших данных в развитии цифровых гуманитарных технологий / СС. Палицына, Т.Л. Михайлова // Будущее технической науки: сборник материалов XIV Международной молодежной научно-практической конференции. 2015. С. 585.
- [4]. Anderson C (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. WIRED. Available from: <http://www.wired.com/2008/06/pb-theory/> (9 февраля 2016).
- [5]. Bowker GC (2005) Memory practices in the sciences. MIT Press Cambridge, MA. See also: Gitelman L (ed.) (2013) Raw Data Is an Oxymoron. Cambridge, MA: MIT Press.
- [6]. Boyd danah and Crawford K (2012) Critical Questions for Big Data. Information, Communication & Society 15(5): 662–679
- [7]. Calo R (2013) Consumer Subject Review Boards: A Thought Experiment. Stanford Law Review Online 66: 97.
- [8]. Crawford K, Gray ML and Miltner K (2014) Critiquing Big Data: Politics, Ethics, Epistemology. International Journal of Communication 8(0): 10.
- [9]. Grimmelmann J (2015) Ethical Culture Clashes in Social Media Research. 2d.laboratorium. Доступно на: <http://2d.laboratorium.net/post/108480841510/ethical-culture-clashes-in-social-media-research> (успешный доступ: 6 ноября 2015).
- [10]. Kramer ADI, Guillory JE and Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences 111(24): 8788–8790.
- [11]. Meyer MN and Chabris CF (2015) Please, Corporations, Experiment on Us. The New York Times, 19 June 2015. Доступно на: <http://www.nytimes.com/2015/06/21/opinion/sunday/please-corporations-experiment-on-us.html> (успешный доступ: 19 октября 2015).
- [12]. Wan Z, Vorobeychik Y, Xia W, et al. (2015) A Game Theoretic Framework for Analyzing Re-Identification Risk. PLOS ONE 10(3): e0120592.