Большие данные в государственном управлении: возможности и угрозы*

ЧАННОВ Сергей Евгеньевич, заведующий кафедрой служебного и трудового права Волжского института управления им. П. А. Столыпина — филиала Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации, доктор юридических наук, профессор

410031, Россия, г. Саратов, ул. Соборная, 23/25

E-mail: sergeychannov@yandex.ru

В статье рассматриваются возможности и перспективы использования больших данных (БД) в государственном управлении на примере Российской Федерации.

Цель исследования — определить, насколько возможно и целесообразно использование БД в государственном управлении в России с учетом накопленного опыта функционирования государственных информационных систем в различных сферах управления, а также уяснить, какие угрозы и риски могут быть с этим связаны.

В процессе исследования использовались диалектический и сравнительный методы, а также методы формальной логики.

Отмечается, что БД при использовании в государственном управлении способны повысить его эффективность. В то же время БД несут и угрозы, которые уже выявлены при использовании данной технологии в коммерческих структурах, что следует учитывать в случае использования БД в государственном управлении. В первую очередь речь идет об угрозе нарушения права на конфиденциальность информации о личности в результате использования БД. Данная технология во многом обесценивает существующие правовые механизмы защиты персональных данных и требует изменения самой концепции охраны информации о личности. С указанной угрозой непосредственно связана и другая, заключающаяся в возможности принятия в результате использования технологии БД неверных управленческих решений, ущемляющих права и свободы граждан. Таким образом, внедрение технологии БД в государственном управлении должно сопровождаться разработкой определенных требований к их применению в отношении граждан и организаций. В частности, должен быть закреплен общий запрет на принятие управленческих решений, порождающих юридические последствия в отношении граждан или организаций либо иным образом затрагивающих их права и законные интересы на основании исключительно автоматизированной обработки любых данных. В ряде ситуаций, возможно, необходимо будет сохранить требование обязательного установления причинно-следственных связей, которые не были выявлены в рамках компьютерного анализа, и доказывания сделанных выводов с помощью логического обоснования.

Ключевые слова: государственное управление, большие данные, управленческие решения, государственные информационные системы, персональные данные, права граждан.

Big Data in Public Administration: Opportunities and Threats

S. E. CHANNOV, head of the Department of service and labor law of the Stolypin Volga Region Institute of Administration — the Branch of the Russian Academy of National Economy and Public Administration under the President of the Russian Federation, doctor of legal sciences, professor

23/25, Sobornaya st., Saratov, Russia, 410031

E-mail: sergeychannov@yandex.ru

The article examines the opportunities and prospects for using big data in public administration on the example of the Russian Federation.

The aim of the study is to determine how far and how much big data can be used in the Russian public administration, taking into account the accumulated experience of the state information systems'

* Публикация подготовлена в рамках поддержанного РФФИ научного проекта № 17-03-00082-ОГН.

functioning in various spheres of management, and also what threats and risks can be associated with this.

The author used general scientific methods: dialectical method, methods of formal logic, and also a comparative method.

Based on the analysis of both Russian and foreign experience the author concludes that big data in public administration can improve its effectiveness. At the same time, it is noted that big data also poses some threats that have already manifested themselves in the use of this technology in commercial structures and that should be considered in the case of the use of big data in public administration. First of all, it is a threat of violation of the right to privacy of personal information as a result of the use of big data. This technology largely devalues the existing legal mechanisms of personal data protection and requires a change in the concept of personal information protection. This threat is directly related to the other, which is the possibility of making incorrect management decisions as a result of the use of big data technology, infringing on the rights and freedoms of citizens. All this leads the author to the conclusion that the introduction of big data technology in public administration should be accompanied by the development of certain requirements for their application to citizens and organizations. In particular, there should be a General ban on management decisions that produce legal consequences for citizens or organizations or otherwise affect their rights and legitimate interests on the basis of exclusively automated processing of any data. In a number of situations, it may be necessary to maintain the requirement to establish causal relationships that have not been identified by computer analysis and to prove the findings by logical reasoning.

Keywords: public administration, big data, management decisions, public information systems, personal data, rights of citizens.

DOI: 10.12737/art_2018_10_11

Всевозрастающее использование технологии больших данных (Big Data, далее — БД) для решения самого широкого круга задач — очевидный тренд большинства стран мира. При этом основной сферой использования БД остается частный сектор. Однако в последние годы идет явная тенденция на использование БД и в государственном управлении¹.

Большие данные определяются специалистами как совокупность инструментов и методов обработки структурированных и неструктурированных данных огромных объемов из различных источников, подверженных постоянным обновлениям, в целях повышения качества принятия управленческих решений, создания новых продуктов и повышения конкурентоспособно-

сти². Другие ученые определяют БД в качестве серии подходов, инструментов и методов обработки структурированных и неструктурированных данных значительных объемов для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети³.

Как видно, в приведенных выше достаточно схожих определениях основной упор делается на методы и способы обработки данных. В то же время на практике термин «большие данные» нередко используется в двух значениях. Так именуется технология, позволяющая анализировать большие объемы информации с высокой скоростью, находя определенные корреляции между ними, а также информация, состав-

¹ См.: Булгакова Е. В., Булгаков В. Г., Акимов В. С. Использование «больших данных» в системе государственного управления: условия, возможности, перспективы // Юридическая наука и практика: Вестник Нижегородской академии МВД России. 2015. № 3. С. 10—14.

² См.: Савельев А. И. Комментарий к Федеральному закону от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и защите информации» (постатейный). М., 2015.

³ См.: Булгакова Е. В., Булгаков В. Г., Акимов В. С. Указ. соч. С. 10—14.

ляющая эти БД (большой набор разнообразных и разнородных данных, поддающихся сегментации по каким-либо признакам). Данные значения отражают две важнейшие составляющие технологии БД, которые можно выразить следующими формулировками:

1) анализу подвергаются все имеющиеся в наличии данные, без каких-либо попыток сделать выборку, причем чем данных больше, тем достоверней получаемые результаты;

2) в рамках указанного анализа алгоритм программы, его проводящей, ищет корреляции между различными данными, не устанавливая их причины (корреляции вместо причинности). Таким образом, выводы, полученные в результате использования технологии БД, как правило, обладают высокой степенью достоверности, однако нередко не могут быть объяснены с логической точки зрения. То есть основное в использовании БД — это извлечение из данных нового смысла, знания, явно в них не содержавшегося⁴.

Как отмечалось, первоначально технология БД была разработана и нашла применение в деятельности частных компаний, преимущественно крупных и обладающих большими массивами данных. Использование БД в государственном управлении — следствие не только уже доказанной эффективности рассматриваемой технологии, но и того факта, что практически любое государство является обладателем огромного массива данных о гражданах (причем нередко не только о своих, но и о чужих). «Государственные органы... по-прежнему да-

дут фору любой частной компании в том, что касается огромного объема управляемых данных. Государственные органы, в отличие от держателей данных в частном секторе, как правило, обязывают людей предоставить информацию, а не убеждают или предлагают что-то взамен. Поэтому они и далее будут собирать и накапливать огромные объемы данных»⁵.

Приведенная цитата верна и применительно к России. Более того, российская система государственного управления характеризуется накоплением даже больших объемов информации о гражданах и организациях, чем это имеет место во многих других государствах. Связано это с массовым внедрением во все сферы государственного управления различного рода информационных систем, служащих для аккумуляирования, хранения и обработки данных.

Разумеется, информационные системы как инструмент государственного управления используются во всех технологически развитых странах мира. Однако в большинстве государств количество их в сфере государственного управления значительно меньше, чем в нашей стране; при этом более широко распространен опыт использования для целей публичного управления частных либо частно-государственных систем⁶.

А вот в России внедрение государственной информационной системы в последние годы зачастую рассматривается как единственный способ повышения эффективности контроля и учета в определенной

⁴ Большие данные в государственном управлении: опыт внедрения (пилотное исследование): материалы круглого стола в Аналитическом центре при Правительстве Российской Федерации. URL: <http://ac.gov.ru/files/content/10087/sorokin-kruglyj-stol-issledovanie-pdf.pdf> (дата обращения: 22.05.2018)

⁵ Майер-Шенберген В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М., 2014. С. 123.

⁶ См.: Амелин Р. В. Государственные и муниципальные информационные системы в российском информационном праве: теоретико-правовой анализ: монография / под ред. С. Е. Чаннова. М., 2018. С. 66—68.

сфере, чуть ли ни панацея от корупционных проявлений. Неудивительно, что только на федеральном уровне количество государственных информационных систем исчисляется сотнями, при этом ежегодно появляется около десятка новых информационных систем, которые начинают рассматриваться как вспомогательный правовой инструмент регулирования общественных отношений⁷. Соответственно, в этих многочисленных государственных информационных системах накапливается все больший объем информации, зачастую более нигде не существующей (т. е. определенная информация существует только в электронном виде и только в базах данных конкретной информационной системы). Такая ситуация является практически идеальной для использования технологий БД.

Это можно только приветствовать, учитывая доказанную высокую эффективность использования БД в коммерческом секторе. Что касается сферы государственного управления, учеными отмечается, что БД «позволят организовать анализ открытых данных для оптимального планирования городских парковок; создать сервис, информирующий автовладельцев о планах городских властей произвести ремонт или уборку улицы; провести сбор информации о состоянии линий электропередачи, дорог, систем водоснабжения в онлайн-режиме; организовывать социально ориентированное движение общественного транспорта при одновременном снижении уровня бюджетных затрат»⁸. Другие специалисты подчеркивают возможности БД в избирательном процессе. «Современный уровень развития информационных технологий позволяет

проанализировать весь объем данных, связанных с участием граждан в избирательных кампаниях, не ограничиваясь статистической выборкой. Это делает возможным определение корреляций, на основании которых можно с недоступной ранее высокой долей вероятности прогнозировать поведение как всего электората, так и отдельных избирателей»⁹.

Подобные возможности уже реализуются в Канаде¹⁰. В США администрация социального обеспечения (SSA) использует анализ БД в виде заявок на инвалидность в целях более быстрого и эффективного принятия решений и выявления мошеннических претензий. Департамент образования США использует передовые методы анализа и обработки естественного языка для обработки правительственных документов и отслеживания изменений в политике, законах и правилах¹¹.

Одним из лидеров в использовании БД в государственном управлении является Пакистан. В стране создана одна из крупнейших в мире база мультибиометрических данных граждан NADRA (National Database & Registration Authority). Уже на первом этапе создания NADRA благодаря анализу собранных данных и их сопоставлению со сведениями, содержащимися в других государственных информационных системах, были выявлены граждане, полу-

⁹ Волков В. Э. Электоральная идентификация: пределы индивидуализации участников избирательного процесса // *Lex Russica*. 2015. № 10. С. 46—56.

¹⁰ См.: McGregor G. The Big Data Election: Political Parties Building Detailed Voter Records // *Ottawa Citizen*. 2014. October 18. URL: <http://goo.gl/7xrTw5> (дата обращения: 22.05.2018).

¹¹ См.: Большие данные в государственном управлении: опыт внедрения (пилотное исследование): материалы круглого стола в Аналитическом центре при Правительстве Российской Федерации.

⁷ См.: Амелин Р. В. Правовой режим государственных информационных систем: монография / под ред. С. Е. Чаннова. М., 2016.

⁸ Булгакова Е. В., Булгаков В. Г., Акимов В. С. Указ. соч. С. 13.

чавшие по две и даже три пенсии, а также те, кто за плату отбывал наказание за настоящих преступников. При выверке списков избирателей выяснилось, что 45% (37 млн) записей в них дублируются — некоторые граждане зарегистрированы более 20 раз, а кто-то уже давно умер. Сопоставив данные об имеющейся у граждан собственности — дорогих автомобилях, домах и многочисленных банковских счетах — со списками налогоплательщиков, создатели NADRA выяснили, что 2,4 млн из них вообще не имеет ИНН, а 1,2 млн хотя и имеют, но не платят налогов.

Интересно, что создание такой системы обошлось государству недорого — оно выделило на ее реализацию небольшой кредит лишь на первом этапе, который потом был успешно возвращен. Система развивалась самостоятельно за счет предоставления платных сервисов заинтересованным коммерческим структурам¹².

В России возможности использования технологии БД на государственном уровне также, безусловно, осознаются¹³. При этом некоторые государственные органы уже достаточно успешно используют эти технологии. Так, российские налоговые органы для обработки представленных в электронном виде деклараций по НДС (предмет проверки), книг покупок и книг продаж, журналов учета выставленных и полученных счетов-фактур используют технологическую базу «АСК НДС-2». В поле зрения налоговиков в первую

очередь попадают те компании, которые превышают критерий 89% вычетов по НДС¹⁴. По информации руководителя ФНС М. Мишустина, применение АСК НДС-2 позволило почти в восемь раз снизить количество заявлений на возмещение НДС высокорисковыми организациями¹⁵.

Внедрение систем, подобных АСК НДС-2, — пример эффективного использования технологий БД в государственном управлении. При этом налоговики не получили право доначислять НДС только на основании разорванной цепочки в АСК НДС-2. Более того, данные этого программного комплекса не могут быть основаниями в суде и могут использоваться только для внутреннего применения. Порядок взыскания налоговой недоимки в бюджет остался прежним: выездная налоговая проверка, выявление фактических обстоятельств сделок, доказательство фиктивного документооборота¹⁶. Правда, высказываются мнения о том, что суды в ряде случаев принимают решения в пользу налоговых органов без достаточных доказательств и ориентируясь только на данные, полученные путем использования АСК НДС-2, однако в целом анализ судебной практики это не подтверждает¹⁷.

При этом не следует забывать, что рассматриваемая технология, перспективы использования которой в России пока только раскрываются, наряду с несомненными преимуществами при использовании в системе

¹² См.: Рудычева Н. Большие данные в госсекторе: опыт Пакистана. URL: http://www.cnews.ru/articles/bolshie_dannye_v_gossektore_opyt_pakistana (дата обращения: 22.05.2018).

¹³ См. об этом: Путин поручил усилить безопасность государственных информационных систем. URL: https://www.m24.ru/articles/tehnologii/28042017/138793?utm_source=CopyBuf (дата обращения: 22.05.2018).

¹⁴ См.: Гин-Барисьявичене К. Ради пополнения казны // ЭЖ-Юрист. 2016. № 6—7. С. 11.

¹⁵ См.: Тимошенко В. А. Совершенствование контроля за исчислением и уплатой НДС на базе «АСК НДС-2» // Право и экономика. 2017. № 11. С. 63.

¹⁶ См.: Как работает программа АСК НДС-2 и способы ее обхода. URL: http://vitvet.com/articles/ask_nds_2/ (дата обращения: 22.05.2018).

¹⁷ См.: АСК НДС-2: судебная практика. URL: <http://www.buhgalteria.ru/article/n163738> (дата обращения: 22.05.2018).

государственного управления может нести и определенные угрозы. Данные риски уже хорошо известны на основе опыта использования БД в коммерческих структурах¹⁸.

Так, в первую очередь можно выделить угрозу нарушения права на конфиденциальность информации о личности в результате использования БД. Данная технология во многом обесценивает существующие правовые механизмы защиты персональных данных. Сейчас уже недостаточно защищать только какие-то определенные данные, ранее считавшиеся критичными для безопасности (имена, номера счетов, идентификационные коды и т. п.), поскольку сегодня современные технологии позволяют без особых проблем идентифицировать личность конкретного лица посредством установления корреляций между несколькими фрагментами данных¹⁹. Попытки придать таким данным анонимную форму (обезличить) также не всегда эффективны ввиду деобезличивающей способности технологий БД²⁰. Что еще хуже: БД позволяют не просто устанавливать личность конкретного человека, но и выявлять информацию о нем, прямо в анализируемых данных отсутствующую (в том числе политические, религиозные, сексуальные взгляды и предпочтения

субъекта, состояние его здоровья и т. п.)²¹.

Помимо этого в настоящее время использование современных цифровых технологий приводит к тому, что привязка данных к конкретному лицу (что, собственно, и определяет их как персональные) в ряде случаев становится уже не столь важной. Как отмечается, «если компания имеет порядка 100 единиц информации обо мне, которые оказывают влияние на то, как она строит свои отношения со мною в цифровой среде, какая разница, знают они мое имя или нет?»²². В современных технических реалиях компании не обязательно знать имя человека для того, чтобы персонализировать свое отношение к нему и предлагать соответствующие товары (услуги)²³.

При этом применение технологии БД противоречит принципу ограничения обработки персональных данных заранее определенными целями, указанному в ст. 5 Федерального закона от 27 июля 2006 г. № 152-ФЗ «О персональных данных»²⁴, поскольку при ее использовании обрабатываются накопленные у организации (или у государ-

¹⁸ См. об этом, например: *Майер-Шенберген В., Кукуер К.* Указ. соч. С. 157—176.

¹⁹ См.: *Narayanan A., Shmatikov V.* Robust De-Anonymization of Large Sparse Datasets // The University of Texas. 2008 IEEE Symposium on Security and Privacy. URL: https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf (дата обращения: 22.05.2018); *Dwork C.* «Differential Privacy». ICALP'06 Proceedings of the 33rd International Conference on Automata, Languages and Programming. Berlin, 2006.

²⁰ См.: *The Myth of Anonymization: Has Big Data Killed Anonymity?* URL: <http://www.kantarhealth.com/docs/white-papers/the-myth-of-anonymization-has-big-data-killed-anonymity.pdf> (дата обращения: 22.05.2018).

²¹ См. об этом: *Mayer J. et al.* Evaluating the Privacy Properties of Telephone Metadata. Stanford University. 1 March 2016. URL: <http://www.pnas.org/content/113/20/5536> (дата обращения: 22.05.2018).

²² *Nissenbaum H. et al.* Privacy, Big Data, and the Public Good: Frameworks for Engagement. Cambridge University Press, 2014. P. 70.

²³ См.: *Савельев А. И.* Электронная коммерция в России и за рубежом: правовое регулирование. 2-е изд. М., 2016.

²⁴ *Булгакова Е. В.* Методы анализа и принципы систематизации данных, используемых при подготовке и принятии юридических решений // Информационное право. 2016. № 4. С. 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.

ства) самые различные персональные данные, собранные ранее для иных целей.

В связи с этим некоторые российские ученые уже призывают к изменению правовых подходов в данной сфере. Так, А. И. Савельев делает вывод, что «в новых технологических реалиях “узкий” подход к понятию персональных данных, “заточенный” под анкетные данные и данные различного рода государственных реестров, уже не способен выполнять функцию эффективного средства защиты частной жизни граждан в сети Интернет», и указывает, что «будущее — за “широким” подходом к определению персональных данных, по крайней мере, пока на смену законодательству о персональных данных не придет принципиально новое регулирование, ориентированное не столько на защиту “анкетных данных” физических лиц, сколько на защиту данных об их поведении в цифровом мире, на основании которых в перспективе будет приниматься множество юридически значимых решений в отношении таких лиц»²⁵.

Если угроза нарушения прав на частную жизнь в связи с постоянным умножением масштабов использования БД существует уже сейчас, то угроза ущемления прав и свобод граждан на основе сделанных путем использования технологии БД предположений об их прошлом и даже будущем поведении имеет пока в большей степени потенциальный характер, по крайней мере в государственном секторе.

Сегодня ряд стран, в частности США, используют БД в целях так называемого прогностического полицейского контроля. В частности, широкую известность получила система “Blue CRUSH” (Crime Reduction Utilizing Statistical History — Снижение преступно-

сти на основе статистических данных), разработанная компанией “IBM”, которая предоставляет полицейским подготовленные на основе имеющейся статистики совершения преступлений сведения о зонах потенциальной угрозы совершения преступления с указанием места (в пределах нескольких кварталов) и времени (в пределах нескольких часов конкретного дня недели). Подобного рода профилактическое прогнозирование привело к снижению уровня преступности в г. Мемфисе на 31%, из которых 15% приходится на тяжкие преступления²⁶. Технические решения, имеющие в своей основе аналитику «больших данных», используются в ряде иных городов США (Нью-Йорк, Сиэтл, Лос-Анджелес и др.), и масштаб их использования возрастает с каждым годом²⁷.

Однако ряд исследователей технологии БД видят в таком подходе и определенные угрозы: «...это скользкий путь. Если на основе анализа больших данных мы сможем прогнозировать возможных преступников, то вряд ли станем довольствоваться профилактикой преступлений. Вероятно, мы захотим наказать потенциальных виновников»²⁸.

Таким образом, суть проблемы заключается в том, что БД позволяют с достаточно высокой степенью вероятности определить будущее противоправное поведение анализируемого лица. Собственно, указанная угроза неразрывно связана с предыдущей, поскольку тоже основана на профайлинге (профили-

²⁶ См.: *Thompson T. Crime Software May Help Police Predict Violent Offences* // *The Guardian*. 2010. July 25. URL: <http://www.theguardian.com/uk/2010/jul/25/police-software-crime-prediction> (дата обращения: 22.05.2018).

²⁷ См.: *Joh E. Policing by Numbers: Big Data and the Fourth Amendment* // *Washington Law Review*. 2014. No. 89:35. P. 35—68.

²⁸ *Майер-Шенберген В., Кукьер К.* Указ. соч. С. 165—166.

²⁵ *Савельев А. И.* Электронная коммерция в России и за рубежом: правовое регулирование. 2-е изд. М., 2016.

ровании), т. е. процедуре выявления характерных ассоциаций в отношении конкретного лица, которые позволяют прогнозировать его поведение. «Профайлинг» — весомое слово. Оно подразумевает не только дискриминацию в отношении определенных групп, но и при неправильном использовании означает «вину по ассоциации». Профайлинг имеет серьезные недостатки»²⁹.

Как отмечалось, данная угроза в указанном «абсолютном» варианте (привлечение к юридической ответственности лица за совершенное еще преступление) является пока чисто теоретической. Однако на практике уже имеют место ситуации наступления негативных последствий для лица на основании проведенного профайлинга с использованием технологии БД и предсказания его будущего поведения. Так, лицу, обратившемуся в банк, может быть отказано в кредите, если анализ покажет, что по совокупности каких-то признаков существует высокая вероятность того, что он этот кредит не выплатит. В другом случае гражданин получает отказ в приеме на работу, поскольку по результатам анализа он признан неблагожелательным кандидатом.

Как отмечают В. Майер-Шенберген и К. Кукьер, подобные методики использовались ранее и при анализе малых данных. Однако технология БД позволяет делать это более тщательно в отношении конкретного лица³⁰. Конечно, решения в конечном счете будут приниматься человеком, а не компьютером, однако маловероятно, что уполномоченное должностное лицо компании, широко использующей анализ БД, без каких-либо веских причин отвергнет результаты такого анализа.

²⁹ Bernard E. Harcourt. *Against Prediction: Profiling, Policing, And Punishing In An Actuarial Age*. University of Chicago Press, 2006.

³⁰ См.: Майер-Шенберген В., Кукьер К. Указ. соч. С. 167.

Рассмотрим в связи с этим следующую, пока еще гипотетическую ситуацию. Российское трудовое законодательство, как и законодательство многих стран мира, закрепляет императивный запрет отказывать в заключении трудового договора женщинам по мотивам, связанным с беременностью или наличием детей (ч. 2 ст. 64 ТК РФ), нарушение которого подкрепляется даже мерами уголовной ответственности (ст. 145 УК РФ). Наличие такого запрета обусловлено тем, что работодатель не заинтересован в трудоустройстве беременной женщины. Однако на практике этот запрет действует далеко не всегда. Связано это с тем, что работодатель в большинстве случаев имеет возможность отказать в приеме на работу беременной женщине, указывая в качестве основания отказа не ее беременность, а какие-либо другие мотивы (например, отсутствие необходимых деловых качеств). Указанное обстоятельство вынуждает многих беременных женщин при устройстве на работу скрывать свою беременность; при этом у работодателя зачастую не имеется законного пути установить этот факт (исключением, например, являются ситуации, когда законом предусмотрен обязательный медицинский осмотр).

Однако современные цифровые технологии БД позволяют узнавать о факте беременности женщины, даже если она об этом явно не сообщала. Так, специалисты американского магазина сниженных цен «Target» на основе анализа запросов и покупок женщин, о беременности которых было точно известно, выявили около двух десятков характерных продуктов, по которым клиенту можно было присвоить оценку «прогнозируемой беременности». Необходимо подчеркнуть, что многие из этих продуктов не имели непосредственного предназначения именно для беременных. Исключительно с помощью корреляций между БД компании удастся опре-

делять даже примерную дату родов!³¹

Данный анализ использовался магазином “Target” исключительно в собственных целях для оптимизации рекламных предложений товаров. Однако, как представляется, компании, обладающие доступом к большим объемам различных данных, вполне могут проводить анализ для определения фактов беременности с целью решения вопроса о трудоустройстве к ним. Не исключено и появление интернет-сервисов, предлагающих провести подобное исследование любым работодателям.

Но подобный анализ, насколько высокой достоверностью он бы ни обладал, все же небезупречен. Вычисленная компьютером 99%-ная вероятность наличия какого-либо факта все же не означает, что он точно имеет место. Технология БД не ориентирована на работу с исключениями, поскольку подобная работа требует уже не установления корреляций, а проведения логического анализа.

В результате, например, женщина, не являющаяся беременной, получит отказ в приеме на работу в силу того, что анализ с использованием БД показал высокую вероятность ее беременности. Что интересно, при массовом распространении подобных подходов и технологий она может получать отказы в трудоустройстве от всех работодателей. При этом она даже не будет знать истинную причину отказов!

Приведенный пример касается ситуации выявления факта, имеющего или не имеющего место в настоящее время (наличие или отсутствие беременности). Однако, как отмечалось выше, БД позволяют строить предположения (как правило, сбывающиеся) и о фактах, которые только могут иметь место в будущем. Такая ситуация, например, имеет место, когда на основе анализа

БД лицу отказывают в предоставлении кредита.

Конечно, здесь можно возразить, что прогноз наступления того или иного события всегда использовался в практике деятельности коммерческих организаций (да и не только коммерческих). Скажем, именно на прогнозах основана вся деятельность страховых компаний, которые производят расчет вероятности страхового случая и на этой основе рассчитывают страховой тариф.

Отличие здесь состоит в том, что технология БД позволяет строить прогнозы на основании корреляций, которые зачастую для человека, принимающего решение, остаются неизвестными и (или) необъяснимыми. В этом плане БД представляют собой своего рода «черный ящик»: мы знаем в первом приближении данные на входе и получаем определенные выводы на выходе, но на основании чего они были сделаны — чаще всего неизвестно (корреляции вместо причинности). А. И. Савельев, анализируя различные потенциальные угрозы использования БД, обоснованно пишет: «Субъект персональных данных и не узнает о том, использовались ли результаты Big Data-аналитики при принятии юридически значимого решения в отношении их, а главное, каковы были исходные данные, которые выступали предметом анализа»³². Проблему он видит в том, что «в настоящее время субъект не может контролировать оборот информации о нем в сети Интернет, а следовательно, иметь реальную возможность требовать ее исключения или корректировки»³³.

Безусловно соглашаясь с его мнением в данном вопросе, в то же время отметим, что проблема глубже. Даже если субъект профайлинга узнал, что решения в отноше-

³¹ Майер-Шенберген В., Кукьер К. Указ. соч. С. 65.

³² Савельев А. И. Электронная коммерция в России и за рубежом: правовое регулирование. М., 2014.

³³ Там же.

нии него принимались на основе БД и выяснил исходные данные, которые выступали предметом анализа, он может никогда не узнать, какие конкретно корреляции были сделаны и почему они привели к этому выводу (принцип черного ящика). Как опровергнуть вывод, основанный неизвестно на чем? Достаточно давно используемый профайлинг на основе малых данных содержит логические цепочки, т. е. устанавливаются причинно-следственные связи. Так, профайлеры, сотрудничающие с правоохранительными органами, основывают свои выводы на вполне определенных, логически объяснимых посылах. «Например, если жертву связали перед убийством, то преступник, вероятно, слаб физически. Сделал из убийства ритуал — скорее всего наслаждался процессом и повторит преступление. Убийца хаотично бил жертву ножом в лицо и грудь, а потом убрал кровавые следы и аккуратно уложил ее на кровать, укрыв с головой, — убийца скорее всего был знаком с жертвой, в какой-то момент потерял контроль, а после испугался и раскаялся»³⁴. При использовании БД причинность заменяется корреляциями.

Как отмечалось, в сфере взаимодействия государства и гражданина ситуации наказания за факты, выявленные на основе только анализа БД, и тем более за предполагаемые в будущем деяния, пока еще остаются по большей части теорией. Так, АСК НДС-2 в настоящее время позволяет на основе анализа деятельности всех налогоплательщиков выявить разрывы в цепочках платежей. Однако такое выявление не является само по себе достаточным для взыскания налоговых недоимок. Оно лишь дает налоговым органам информацию о том, какие организа-

ции, по всей видимости, допустили налоговые нарушения. Дальнейшее расследование идет в так называемом ручном режиме.

В рассматриваемом случае связь между анализом БД и действиями организаций-налогоплательщиков достаточно прямая и очевидная. Однако ситуация вполне может измениться при усложнении подобных информационных систем, используемых в государственном управлении. Если количество данных будет очень велико, а корреляции многочисленны — понять причину сигнала системы об отклонении поведения контролируемого субъекта от нормального будет уже не так просто. И тогда вполне может появиться соблазн принять решение непосредственно на основе вывода системы, причем чем точнее будут выводы системы, сделанные ранее, тем этот соблазн будет выше. А здесь уже недалеко и до следующего шага — попыток предсказать поведение контролируемого субъекта на основе анализа БД и привлечь его к ответственности еще до того, как он совершит правонарушение.

Угроза использования наказаний без преступления при использовании БД в государственном управлении возможна и в других сферах. В частности, это может иметь место и при взаимодействии частных компаний с гражданами, осуществляемом в рамках строго регулируемых государством сферах.

Прогноз наступления страхового случая — важнейшая составляющая деятельности страховых компаний. При этом многие страховые компании уже используют технологию БД для расчета страховых тарифов. Естественно, вполне может сложиться ситуация, когда для определенных клиентов индивидуальный тариф оказывается выше, чем для других категорий, причем причины этого объяснить не всегда возможно (корреляции вместо причинности). Однако в частной сфере

³⁴ Российские профайлеры. Психологи на службе у силовиков: зачем они нужны и есть ли от них польза. URL: <https://meduza.io/feature/2016/11/18/rossiyskie-profaylery> (дата обращения: 22.05.2018).

клиент, даже если он лишен возможности опровергнуть сделанный в отношении него вывод, всегда может отказаться от заключения договора страхования (или отказать это сделать с конкретной компанией). Иное дело — когда оно является обязательным. Такая ситуация может сложиться, если, например, БД станут использоваться при расчете тарифов для заключения договора ОСАГО.

Вышеизложенное приводит нас к выводу, что внедрение технологии БД в государственном управлении должно сопровождаться разработкой определенных требований к их применению в отношении граждан и организаций. Здесь уместно напомнить, что Федеральный закон «О персональных данных» содержит общий запрет на принятие на основании исключительно автоматизированной обработки персональных данных решений, порождающих юридические последствия в отношении субъекта персональных данных или иным образом затрагивающих его права и законные интересы (ст. 16). Однако БД отнюдь не всегда являются персональными, поэтому на основе анализа БД принятие подобных решений вполне законно.

Логичной выглядит формулировка подобного запрета уже в отношении не только персональных, но и всех данных, подвергающихся автоматизированной обработке. Правда, к сожалению, это тоже не будет ис-

черпывающим решением проблемы. Принцип «черного ящика», лежащий в основе технологии БД, приведет к тому, что субъект, принимающий управленческое решение, зачастую не сможет никак обосновать вывод, сделанный программой (информационной системой), и решение его скорее всего будет иметь формальный характер. Таким образом, при внешнем соблюдении буквы закона дух его все равно будет нарушаться.

Вряд ли сейчас можно предложить универсальный подход для данной проблемы. Возможно, в будущем придется искать баланс между решением, принимаемым полностью в автоматическом режиме, и решением, принимаемым человеком — по-разному для разных сфер управления, а также с учетом значимости юридических последствий указанных решений. В ряде ситуаций, возможно, необходимо будет сохранить требование обязательного установления причинно-следственных связей, которые не были выявлены в рамках компьютерного анализа, и доказывания сделанных выводов с помощью логического обоснования (подобно тому, как система «АСК НДС-2» подсказывает — где искать, но не освобождает от необходимости доказывания факта правонарушения человеком). Это, конечно, в определенной степени обесценивает преимущества БД, однако в противном случае мы получаем слишком большие риски.

Библиографический список

- Bernard E. Harcourt. *Against Prediction: Profiling, Policing, And Punishing In An Actuarial Age*. University of Chicago Press, 2006.
- Dwork C. "Differential Privacy". ICALP'06 Proceedings of the 33rd International Conference on Automata, Languages and Programming, Berlin, 2006.
- Joh E. *Policing by Numbers: Big Data and the Fourth Amendment* // Washington Law Review. 2014. No. 89:35.
- Mayer J. et al. *Evaluating the Privacy Properties of Telephone Metadata*. Stanford University. 1 March 2016. URL: <http://www.pnas.org/content/113/20/5536>.
- McGregor G. *The Big Data Election: Political Parties Building Detailed Voter Records* // Ottawa Citizen. 2014. October 18. URL: <http://goo.gl/7xrTw5>.
- Narayanan A., Shmatikov V. *Robust De-Anonymization of Large Sparse Datasets* // The University of Texas. 2008 IEEE Symposium on Security and Privacy. URL: https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.

Nissenbaum H. et al. Privacy, Big Data, and the Public Good: Frameworks for Engagement. Cambridge University Press, 2014.

The Myth of Anonymization: Has Big Data Killed Anonymity? URL: <http://www.kantarhealth.com/docs/white-papers/the-myth-of-anonymization-has-big-data-killed-anonymity-.pdf>.

Thompson T. Crime Software May Help Police Predict Violent Offences // The Guardian. 2010. July 25. URL: <http://www.theguardian.com/uk/2010/jul/25/police-software-crime-prediction>.

Амелин Р. В. Государственные и муниципальные информационные системы в российском информационном праве: теоретико-правовой анализ: монография / под ред. С. Е. Чаннова. М., 2018.

Амелин Р. В. Правовой режим государственных информационных систем: монография / под ред. С. Е. Чаннова. М., 2016.

Булгакова Е. В., Булгаков В. Г., Акимов В. С. Использование «больших данных» в системе государственного управления: условия, возможности, перспективы // Юридическая наука и практика: Вестник Нижегородской академии МВД России. 2015. № 3.

Булгакова Е. В. Методы анализа и принципы систематизации данных, используемых при подготовке и принятии юридических решений // Информационное право. 2016. № 4.

Волков В. Э. Электоральная идентификация: пределы индивидуализации участников избирательного процесса // Lex Russica. 2015. № 10.

Гин-Барисявичене К. Ради пополнения казны // ЭЖ-Юрист. 2016. № 6-7.

Как работает программа АСК НДС-2 и способы ее обхода. URL: http://vitvet.com/articles/ask_nds_2/.

Майер-Шенберген В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М., 2014.

Путин поручил усилить безопасность государственных информационных систем. URL: https://www.m24.ru/articles/tehnologii/28042017/138793?utm_source=CopyBuf.

Российские профайлеры. Психологи на службе у силовиков: зачем они нужны и есть ли от них польза. URL: <https://meduza.io/feature/2016/11/18/rossiyskie-profaylery> (дата обращения: 22.05.2018).

Рудычева Н. Большие данные в госсекторе: опыт Пакистана. URL: http://www.cnews.ru/articles/bolshie_dannye_v_gossektore_opyt_pakistana.

Савельев А. И. Комментарий к Федеральному закону от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и защите информации» (постатейный). М., 2015.

Савельев А. И. Проблемы применения законодательства о персональных данных в эпоху «Больших данных» (BIG DATA) // Право. Журнал Высшей школы экономики. 2015. № 1.

Савельев А. И. Электронная коммерция в России и за рубежом: правовое регулирование. 2-е изд. М., 2016.

Савельев А. И. Электронная коммерция в России и за рубежом: правовое регулирование. М., 2014.

Тимошенко В. А. Совершенствование контроля за исчислением и уплатой НДС на базе «АСК НДС-2» // Право и экономика. 2017. № 11.