

*Алешко Д. А., студент, Федеральное Государственное Бюджетное
Образовательное Учреждение Высшего Образования "Саратовская
государственная юридическая академия"*

2 курс Институт правоохранительной деятельности, Россия, г. Саратов

О ВОПРОСЕ РЕАЛИЗАЦИИ ПРАВА НА НЕПРИКОСНОВЕННОСТЬ ЧАСТНОЙ ЖИЗНИ ГРАЖДАН В ЭПОХУ ТЕХНОЛОГИИ "BIG DATA"

Аннотация: В последнее время электронное образование во всем мире стремительно развивается и основной проблемой становится своевременное обеспечение учащихся качественной учебной информацией. Эту задачу невозможно решить без анализа большого потока информации, поступающего в информационную среду электронного образования от участников образовательного процесса – студентов, преподавателей, администрации и т.д. В этой среде существует большое количество различных типов данных, как структурированных, так и неструктурированных, обработку которых трудно осуществить традиционными статистическими методами. Целью исследования является показать, что для разработки и внедрения успешных систем электронного обучения необходимо использовать новые технологии, которые позволили бы хранить и обрабатывать большие потоки данных. Для хранения больших данных требуется большой объем дисковой памяти. Показано, что для решения этой проблемы эффективно использовать кластерную технологию NAS (Network Area Storage), позволяющая хранить информацию учебных заведений на NAS – серверах и иметь к ним общий доступ из Интернета. Для обработки и персонализации Больших Данных в среде электронного образования предлагается использовать технологии MapReduce, Hadoop, NoSQL и другие. В статье приводятся примеры использования этих технологий в облачной среде. Эти технологии в электронном образовании позволяют

достигнуть гибкости, масштабируемости, доступности, качества обслуживания, безопасности, конфиденциальности и простоты использования учебной информации.

В статье показано, что на сегодняшний день разработано множество программных приложений для интеллектуального анализа больших данных. Эти программные продукты можно использовать для классификации, кластеризации, регрессионного и сетевого анализа учебной информации. Применение этих методов в электронном образовании позволит педагогам своевременно получать информацию об обучающихся, оперативно реагировать на любые изменения процесса обучения, своевременно вносить изменения в учебный контент. Полученные результаты исследования предлагается использовать для выработки рекомендаций при создании электронных курсов в высших и средних учебных заведениях Азербайджана.

Ключевые слова: Обезличивание персональных данных, кластер, консалтинг, нейронная сеть, интерфейс, информационные технологии.

Annotation: Recently, e-education has been rapidly developing all over the world, and the main problem is the timely provision of high-quality educational information to students. This problem cannot be solved without analyzing the large flow of information coming to the information environment of e-education from participants in the educational process-students, teachers, administration, etc. In this environment, there are a large number of different types of data, both structured and unstructured, which are difficult to process using traditional statistical methods. The aim of the study is to show that in order to develop and implement successful e-learning systems, it is necessary to use new technologies that would allow storing and processing large data streams. Storing large data requires a large amount of disk memory. It is shown that to solve this problem, it is effective to use the cluster NAS technology (Network Area Storage), which allows storing information from educational institutions on NAS servers and having shared access to them from the Internet. It is proposed to use MapReduce, Hadoop, NoSQL and other technologies

for processing and personalization of Big Data in the e-education environment. This article provides examples of using these technologies in a cloud environment. These technologies in e-education allow you to achieve flexibility, scalability, availability, quality of service, security, confidentiality and ease of use of educational information.

The article shows that many software applications for big data mining have been developed to date. These software products can be used for classification, clustering, regression, and network analysis of educational information. The use of these methods in e-education will allow teachers to receive timely information about students, promptly respond to any changes in the learning process, and make timely changes to the educational content. The results of the study are proposed to be used to develop recommendations for creating e-courses in higher and secondary educational institutions of Azerbaijan.

Keywords: Depersonalization of personal data, cluster, consulting, neural network, interface, information technologies.

Вопросы применения информационных технологий важны для любого государства, поскольку они предполагают конкурентоспособность стран на мировой арене [3]. Одной из таких передовых технологий являются Большие данные, которые способны выполнять множество функций (оплаты, товара, методики) в различных сферах жизнедеятельности [4].

В данной работе рассмотрены проблемные вопросы, связанные с применением технологии «Больших данных», а также представлены положительные и отрицательные факторы влияния Big Data на личную жизнь граждан. Подводя к выводу о том, что ввиду отсутствия закона о регулировании «больших данных» основным препятствием для злоупотребления частной информацией граждан является Закон о персональных данных, однако легитимное согласие на обработку персональных данных, их обезличивание, закрепленные в законе, не обеспечивают должной защиты личных сведений от использования третьими лицами, поэтому

требуется серьезная проработка вопроса о введении уголовной ответственности за использование личной информации физических лиц с использованием технологии «Больших данных», повлекшее нарушение неприкосновенности частной жизни.

Впервые название «большие данные» возникло в печати в 2008 году, иногда соредактор журнала Nature Клиффорд Л. Линч выпустил заметку про формирование будущего науки с поддержкой технологий службы с большим количеством данных. До 2009 года данный термин рассматривался исключительно с позиции научного анализа, но после выхода вновь много заметок пресса стала свободно пользоваться понятие Big Data – и продолжает употреблять его в настоящее время.

В 2010 году стали являться первоначальные попытки решить нарастающую вопрос больших данных. Были выпущены программные продукты, воздействие которых имелось направлено на то, дабы минимизировать опасности около употребление больших информационных массивов. К 2011 году большими данными увлеклись подобные крупные компании, как Microsoft, Oracle, EMC и IBM – они стали первыми пользоваться наработки Big data в своих стратегиях развития, притом достаточно успешно. Университеты приступили проводить исследование больших данных в качестве некоторого объекта уже в 2013 году ныне проблемами в это сфере занимаются не только науки о данных, но и инженерии совместно с вычислительными предметами.

Изначально в совокупность подходов и технологий включались средства массово-параллельной обработки неопределённо структурированных данных, такие как СУБД NoSQL, алгоритмы MapReduce и средства проекта Hadoop. В дальнейшем к технологиям больших данных стали относить и другие решения, обеспечивающие сходные по характеристикам возможности по обработке сверхбольших массивов данных, а также некоторые аппаратные средства. MapReduce — модель распределённых параллельных вычислений в компьютерных кластерах, представленная компанией Google. Согласно этой

модели приложение разделяется на большое количество одинаковых элементарных заданий, выполняемых на узлах кластера и затем естественным образом сводимых в конечный результат.

NoSQL (от англ. Not Only SQL, не только SQL) — общий термин для различных нереляционных баз данных и хранилищ, не обозначает какую-либо одну конкретную технологию или продукт. Обычные реляционные базы данных хорошо подходят для достаточно быстрых и однотипных запросов, а на сложных и гибко построенных запросах, характерных для больших данных, нагрузка превышает разумные пределы и использование СУБД становится неэффективным.

Hadoop — свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов. Считается одной из основополагающих технологий больших данных.

R — язык программирования для статистической обработки данных и работы с графикой. Широко используется для анализа данных и фактически стал стандартом для статистических программ.

Аппаратные решения. Корпорации Teradata, EMC и др. предлагают аппаратно-программные комплексы, предназначенные для обработки больших данных. Эти комплексы поставляются как готовые к установке телекоммуникационные шкафы, содержащие кластер серверов и управляющее программное обеспечение для массово-параллельной обработки. Сюда также иногда относят аппаратные решения для аналитической обработки в оперативной памяти, в частности, аппаратно-программные комплексы Hana компании SAP и комплекс Exalytics компании Oracle, несмотря на то, что такая обработка изначально не является массово-параллельной, а объёмы оперативной памяти одного узла ограничиваются несколькими терабайтами.

Консалтинговая компания McKinsey, кроме рассматриваемых большинством аналитиков технологий NoSQL, MapReduce, Hadoop, R, включает в контекст применимости для обработки больших данных также

технологии Business Intelligence и реляционные системы управления базами данных с поддержкой языка SQL.

Международная консалтинговая компания McKinsey, специализирующаяся на решении задач, связанных со стратегическим управлением, выделяет 11 методов и техник анализа, применимых к большим данным.

Методы класса Data Mining (добыча данных, интеллектуальный анализ данных, глубинный анализ данных) — совокупность методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных знаний, необходимых для принятия решений. К таким методам, в частности, относятся обучение ассоциативным правилам (association rule learning), классификация (разбиение на категории), кластерный анализ, регрессионный анализ, обнаружение и анализ отклонений и др. Краудсорсинг — классификация и обогащение данных силами широкого, неопределённого круга лиц, выполняющих эту работу без вступления в трудовые отношения.

Смешение и интеграция данных (data fusion and integration) — набор техник, позволяющих интегрировать разнородные данные из разнообразных источников с целью проведения глубинного анализа (например, цифровая обработка сигналов, обработка естественного языка, включая тональный анализ, и др.).

Машинное обучение, включая обучение с учителем и без учителя — использование моделей, построенных на базе статистического анализа или машинного обучения для получения комплексных прогнозов на основе базовых моделей.

Искусственные нейронные сеть, сетевой анализ, оптимизация, в том числе генетические алгоритмы (genetic algorithm — эвристические алгоритмы поиска, используемые для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомым параметров с использованием механизмов, аналогичных естественному отбору в природе).

Сегодня компании нацелены на улучшение взаимодействия сотрудников на предприятии, максимизацию доходов от существующих активов и уменьшение операционных расходов. Поэтому возможность подключения к сети станет основополагающей для нового поколения производителей, а технологии для анализа «больших данных» (big data) — еще более востребованными.

Промышленные предприятия постоянно сталкиваются с необходимостью обработки данных, которые в большом объеме поступают с заводов или других объектов. Компании собирают триллионы байт информации о своих клиентах, поставщиках и производственных операциях. Согласно исследованию консалтинговой компании IDC, к 2020 г. объем ценных данных увеличится в два раза, и 60% всех собранных данных можно будет использовать в процессе принятия решений.

На производстве ценная информация зачастую рассредоточена по нескольким базам данных, приложениям корпоративного уровня и операционным системам и представлена в разных форматах. При отсутствии единой целенаправленной стратегии обработки информации сотрудникам завода и администрации сложно плодотворно взаимодействовать друг с другом и принимать эффективные решения.

Как управлять большими массивами данных? Обработка огромного объема неструктурированных данных, их фильтрация и правильная интерпретация становятся первостепенными задачами для предприятий. Важную роль в их решении играет семантическое представление информации, в частности ее визуализация с помощью построения иерархической модели процессов и производственных установок. Данные должны быть представлены в понятном для пользователей виде и в контексте стоящих перед ними задач. Этой цели служит, например, Uniformance PHD — один из компонентов пакета Uniformance Suite компании Honeywell, созданный для сбора и хранения данных о технологических процессах и событиях в реальном времени.

Решения, обеспечивающие непрерывный контроль рабочих показателей, позволяют отправлять уведомления, формировать задания, запускать рабочие процессы и способствовать решению проблем и выявлению новых возможностей в режиме реального времени. Они также могут включать в себя многофункциональные информационные панели. Данные обрабатываются и доставляются во все отделы предприятия, что позволяет наладить взаимодействие между сотрудниками разных подразделений и способствует принятию правильных решений. Кроме того, компании получают возможность заменить быстро устаревающую бумажную документацию на электронный документооборот, а также собирать и сохранять экспертные знания специалистов.

Интеграция и визуализация данных жизненно необходимы для любого промышленного предприятия. Очень важно предвосхищать возможные события, сотрудничать и действовать, основываясь на надежной информации. Принцип работы технологий для управления «большими данными» довольно прост: информация непрерывно собирается из массива датчиков, исполнительных механизмов, контроллеров и человеко-машинных интерфейсов, затем сохраняется и воспроизводится. Своевременное предоставление оперативных данных как на производственном уровне, так и в масштабах всего предприятия позволяет в режиме реального времени обеспечить оперативный контроль и управление технологической установкой или зоной.

Чтобы преобразовать информацию в знания, необходимо не только собирать и структурировать большие массивы производственных данных — нужно также обеспечить современные возможности архивации. Кроме того, использование вычисляемых тегов позволяет применять знания из области инжиниринга и бизнеса для обработки текущих и исторических данных, а встроенные преобразования инженерных единиц — просматривать данные в понятном для пользователей формате. Необходимо также учитывать, что зачастую данные поступают не с одного, а с нескольких объектов, которые

могут быть географически широко распределены. Таким образом, информация, приходящая из разных источников, должна быть консолидирована в единой базе данных, на едином сервере. После установки сервера можно добавлять коллекторы данных по мере роста потребностей конкретного предприятия.

Предотвращение простоев производства. За счет внедрения передовых технологий руководители предприятий могут не только принимать более эффективные решения, но также совершенствовать производственные процессы и решать ключевые бизнес-задачи. Например, получение своевременной и важной информации позволяет увеличить производительность и повысить уровень безопасности на предприятии. Как показало исследование, проведенное Honeywell совместно с KRC Research в прошлом году, незапланированные простои считаются основной угрозой для роста доходов. 42% респондентов (руководителей промышленных предприятий) признались, что эксплуатируют оборудование с большей нагрузкой, чем следует.

Чтобы минимизировать потери от простоев, предприятиям необходимо задуматься о внедрении технологий, которые бы распознавали возможные проблемы и помогали их предотвратить. Непрерывный мониторинг ключевых производственных параметров как раз позволяет операторам своевременно выявить проблему и принять соответствующие меры для ее решения. Например, информация, которая в режиме реального времени поступает из встроенных в трубы сенсоров, в сочетании с данными из других производственных объектов дает возможность распознать скрытую коррозию и предотвратить аварийную ситуацию. На решение подобных задач направлен продукт Uniformance Asset Sentinel в составе аналитической платформы Uniformance Suite.

Эта система осуществляет непрерывный контроль работоспособности оборудования и технологических процессов, позволяя промышленным предприятиям прогнозировать и предотвращать простои и снижение производительности. Возможно это благодаря двум особенностям. Во-первых, система может собирать и консолидировать большие объемы данных, а затем

их анализировать. Она объединяет архивные и визуальные данные, а также информацию, поступающую из распределенной системы управления (PCY) и программируемых логических контроллеров (ПЛК), интеллектуальных средств измерения и из систем мониторинга вибрации. Во-вторых, система использует шаблоны оборудования. Пользователь определяет список переменных и другие конфигурации для конкретного типа оборудования. Каждый тип оборудования достаточно сконфигурировать только один раз, чтобы затем многократно применять при создании модели актива в иерархической структуре предприятия. На этой стадии производится ассоциация объекта с конкретными тегами, поэтому, если необходимо внести какие-либо изменения в конфигурацию, нужно просто поменять шаблон, после чего все его экземпляры будут обновлены автоматически. Кроме того, если пользователи хотят проверить расчеты, им необязательно знать имена тегов (переменных), можно только выбрать нужный актив (оборудование) и интересующие параметры для просмотра. В результате оператор получает всю информацию о состоянии оборудования, включая данные о его эксплуатационных характеристиках и механическом состоянии, а также о вспомогательных устройствах и интеллектуальных средствах измерения.

Внедрение таких технологий позволяет предприятиям из разных отраслей промышленности добиться определенных преимуществ:

- увеличить эффективность использования производственных активов на 10% за счет сокращения количества незапланированных простоев;
- снизить затраты на техническое обслуживание на 10%, усовершенствовав процедуры прогнозирования и предотвращения катастрофических отказов оборудования и выявляя неэффективные операции;
- повысить производительность на 10%;
- сократить эксплуатационные расходы на 10% за счет более эффективного использования энергии.

Сегодня бизнес производит гигантские объемы данных, из анализа и управления которыми генерируются новые знания и эффективные бизнес-

решения. Институт статистических исследований и экономики знаний НИУ ВШЭ в рамках пилотного обследования оценил уровень использования технологий сбора, обработки и анализа больших данных в организациях российской промышленности и ИТ-отрасли и готовность к внедрению таких технологий в будущем. Исследование также показало, в каких бизнес-процессах большие данные применяются чаще всего и насколько востребованы решения российских разработчиков на внутреннем и зарубежном рынке.

Оценки востребованности технологий сбора, обработки и анализа больших данных основаны на результатах опроса 2019 года руководителей 1784 организаций (1163 организации промышленности и 621 ИТ-компания) из 30 регионов России, представляющих все федеральные округа.

Технологии сбора, обработки, анализа больших данных (большие данные) — технологии автоматизированного сбора, обработки, хранения и использования структурированных и неструктурированных массивов информации, характеризующихся значительным объемом и быстрой скоростью изменений.

Чуть более трети (36%) опрошенных руководителей промышленных предприятий сообщили о внедрении технологий сбора, обработки и анализа больших данных. Планы по их дальнейшему использованию имеет почти каждая пятая организация (18,5%): в 2020 г. — 5%, в течение ближайших пяти лет — 14,2%. Наиболее востребованы эти технологии в добывающей промышленности: их применяют 39,8% организаций, еще 28% — планируют внедрить в будущем.

Свыше четверти (28,5%) организаций промышленности используют большие данные в бизнес-процессах. Об их применении в управлении и администрировании (бухгалтерия, финансы, кадры и др.) сообщили 20% респондентов, при производстве продукции и услуг — 12%. Не более 10% руководителей отметили использование таких технологий в других бизнес-процессах: в обработке заказов, транспортировке и хранении продукции, комплектующих — 9,7%, маркетинге и продажах — 7,4%, обслуживании

информационно-телекоммуникационных систем — 5%, разработке продукции, услуг, бизнес-процессов — 4%.

Разработка решений на основе технологий сбора, обработки и анализа больших данных ведется в 44,6% ИТ-компаний, еще 6,9% планируют заниматься этим в ближайшие пять лет. Немногим более 30% ИТ-компаний разрабатывали данные технологии для российского бизнеса. Заказы для органов власти и других отечественных организаций поступали гораздо реже — 10,6 и 11% соответственно. Около 8% руководителей ИТ-компаний сообщили о выполнении заказов для зарубежных компаний, 5,5% — для физических лиц.

Структура источников больших данных, как внутренних, так и внешних, в организациях промышленности и ИТ-компаниях практически одинакова. Чаще всего представители обеих индустрий указывали на использование собственных больших данных. Немногим менее половины организаций промышленности (43,7%) и ИТ-компаний (43,6%) применяли в аналитических целях сведения учетных систем организации, таких как ERP, CRM, SCM, HRIS. Обработку данных корпоративного портала (веб-сайта) проводили четверть (27,8%) организаций промышленности и более чем 40% ИТ-компаний. Активность применения данных геолокации, получаемых, в том числе, с портативных устройств, и сведений, считываемых с цифровых датчиков или радиочастотных меток различного оборудования организации (интернет вещей), сопоставима: в промышленных организациях — 12,2 и 13,4% соответственно, ИТ-компаниях — 15 и 17,4%.

Что касается внешних источников, информацию из социальных сетей и других информационных ресурсов используют 20,8% организаций промышленности, операторов связи — 21,1%. Структура востребованности внешних данных ИТ-компаниями несколько иная — 31,2 и 24,8% соответственно.

«Большие данные» и законодательство персональных данных

Регулирование персональных данных в России осуществляется с применением ряда нормативно-правовых актов:

- №152-ФЗ «О персональных данных» [1];
- №149-ФЗ «Об информации, информационных технологиях и о защите информации» [2];
- ТК РФ Статья 86. Общие требования при обработке персональных данных работника и гарантии их защиты;
- Конвенция Совета Европы о защите физических лиц при автоматизированной обработке персональных данных;
- №125-ФЗ «Об архивном деле в Российской Федерации»;
- и множества других, а если происходит обработка данных жителей Европейского Союза, то ещё и GDPR, который имеет экстерриториальное действие.

Но подробно мы разберем его ниже.

Впрочем, для юристов не так страшны федеральные законы, конвенции и кодексы, как приказы ФСБ, ФСТЭК и Роскомнадзора. Вот только некоторые из них: Приказ Федеральной службы безопасности Российской Федерации от 10 июля 2014 г. №378; Приказ ФСТЭК России от 23 марта 2017 г. №49; Приказ Роскомнадзора от 30 октября 2018 г. №159.

И такие приказы выходят с завидной регулярностью, что для крупных операторов, данных (телеком, банки, интернет-гиганты) является серьезной юридической нагрузкой.

Что такое большие данные и как они связаны с персональными? Российское законодательство не даёт большим данным чёткого определения. По сути это массив структурированных и неструктурированных данных (не только персональных), которые динамически меняются, обрабатываются и анализируются программными инструментами с целью выявления новых связей для принятия различных решений.

Буква закона №152-ФЗ «О персональных данных» определяет персональные данные в качестве любой информации, относящейся к прямо или

косвенно определенному или определяемому физическому лицу (субъекту персональных данных) [1]. Это значит, что фактически любая информация, которая достоверно может определить физическое лицо, к которому эта информация относится, является персональными данными.

В 2018 году группа депутатов внесла на рассмотрение в Госдуму поправки в №149-ФЗ, где вводились такие определения, как «большие пользовательские данные», «оператор больших пользовательских данных», «обработка больших пользовательских данных» и другие [2]. Летом 2019-го журнал Forbes писал об основных проблемах в поправках к федеральному закону об информации.

В итоге никаких поправок так внесено и не было, а законопроект всё ещё находится на рассмотрении.

Большие данные состоят из персональных данных? Персональные данные — это важная составляющая больших данных, но далеко не единственная. Большие данные могут добывать и из других источников: с помощью геолокационных систем и промышленного оборудования. Например, компания Monsanto (сейчас она принадлежит Bayer) долгое время специализировалась на сборе и обработке больших сельскохозяйственных данных. Корпорация собирала показатели уровня влажности, температуры, урожайности и десятков тысяч других фактов, чтобы предсказывать заинтересованным фермерам урожайность их земель в очередном году и давать рекомендации по её повышению. Даже для самого строгого государственного чиновника очевидно, что данные о количестве колосьев пшеницы, собранных в Техасе в 1986 году, нельзя отнести к персональным.

Операторы больших данных обезличивают персональные данные, обрабатывают их в агрегированном виде и получают статистические или демографические данные. Они являются производными из персональных данных, но юридически не классифицируются как таковые, потому что прямо или косвенно не раскрывают личность субъекта, к которому эти данные относятся. Однако при объединении или связи агрегированных данных с

персональными данными появляется возможность идентифицировать физическое лицо. Поэтому такие комбинированные данные будут обрабатываться как персональные.

При этом агрегирование — это лишь один из способов обезличивания персональных данных. Согласно определению, данного в №152-ФЗ, обезличивание — способ обработки персональных данных, в результате которого в обработанных персональных данных нельзя идентифицировать физическое лицо, которому эти данные принадлежат [1]. По сути происходит скрывание, изменение или удаление персональных идентификаторов из набора данных. Конкретные способы и порядок обезличивания устанавливается приказом Роскомнадзора от 5 сентября 2013 г. №996.

На данный момент закон в явной форме не описывает последствия обезличивания данных. Единственное право, которое даёт закон оператору в отношении обработки персональных данных, содержится в п. 9 ч. 1 ст. 6 №152-ФЗ [1]. Оно разрешает использовать обезличенные данные в научных и исследовательских целях. Коммерческое использование данных невозможно без их передачи другим компаниям, что в силу текущих формулировок закона достаточно рискованно.

Также возникают проблемы и в социально-значимых проектах, где не подразумевается коммерческое использование данных. Так, МегаФон совместно с «Лиза Алерт» в марте 2019 года запустил специальную платформу «МегаФон.Поиск» для поиска пропавших детей и взрослых с использованием технологии анализа больших данных, в ходе тестирования которой с помощью платформы удалось найти более 60 человек. Одна из причин использования платформы на базе больших данных, а не просто передача геолокационных данных пропавшего, — ограничения №152-ФЗ «О персональных данных» [1].

Я написала отдельный материал про то, как большие данные помогают искать пропавших.

Идентификаторы и регуляторы

Перечень данных, который можно отнести к персональным, не является закрытым — это вытекает из определения персональных данных, которое дается в №152-ФЗ. Поэтому любой набор информации, позволяющий определить или идентифицировать вас среди множества других людей, относится к персональным. Даже если информация не позволяет точно идентифицировать физическое лицо, но помогает значительно ограничить круг тех людей, к которым данные могут относиться, она является «персональной».

Таких идентифицирующих данных огромное количество, и к ним можно отнести фамилию, имя, отчество, дату рождения, место рождения, возраст, фотографию, ссылку на профиль в социальных сетях и другие.

Например, если вы кому-то сказали вашу фамилию, имя, отчество и дату рождения, то вас с высокой точностью можно будет определить, как конкретную личность. Однако, если мы уберем из набора такие вводные данные, как фамилию или дату рождения, то понять, о каком именно человеке идет речь, будет невозможно. Верно и обратное: безобидные данные, такие как показания электрического счётчика, при их обогащении данными, относящимися к физическим лицам, становятся персональными.

Если говорить о номере мобильного телефона и электронной почте, то представители государственных органов считают, что их использование в качестве персональных данных возможно в отдельных случаях. Например, если имеются дополнительные данные, которые позволяют однозначно соотнести абонентский номер или адрес электронной почты к физическому лицу. Абонентский номер и адрес электронной почты точно не относятся к персональным данным, если договор с оператором был заключен на юридическое лицо. Также почта может являться рабочей или вовсе не содержать никакой идентифицирующей информации.

Бывают и более сложные категории данных. До сих пор нет однозначной позиции о том, является ли IP-адрес персональными данными и обрабатывают ли персональные данные службы Google Analytics и Яндекс.Метрика — позиция регуляторов очень подвижна. В декабре 2018-го Роскомнадзор

заблокировал сайт «Умного голосования», сославшись на тот факт, что владельцы ресурса не уведомили пользователей об использовании счётчиков. В то же время ряд судов явно выражал позицию о том, что IP-адреса к персональным данным не относятся. Да и сам Роскомнадзор не спешит блокировать и штрафовать государственные органы, на сайтах которых без каких-либо уведомлений используются аналогичные инструменты.

Какой вывод из этого можно сделать? Совокупность данных, которые необходимы и достаточны для идентификации физического лица, и понимается под персональными данными. При этом финальное слово всё равно остаётся за судами и контролирующими органами.

Ассоциация больших данных

Поскольку российское законодательство несовершенно и даже между регуляторами возникают разночтения, рынок сам пытается установить для себя общие правила по использованию данных — по крайней мере среди крупных игроков. Банки, телеком-операторы и интернет-гиганты создали Ассоциацию больших данных, в которую также входит и МегаФон. Ассоциация стимулирует своих членов относиться к данным ответственно, не использовать для бизнеса данные с серых рынков, формировать этические нормы в отношении сбора и обработки больших данных для защиты интересов пользователей.

На Неделе Российского Интернета крупнейшие банки, телеком-операторы и крупнейшие интернет-компании подписали кодекс этики использования данных. По словам президента АБД Анны Серебряниковой, целью создания кодекса является развитие «осознанности и добросовестности бизнеса», а также налаживание диалога с государством, и защита интересов граждан в вопросе обращения с данными. Кодекс не несет жесткого административного характера, а следить за его выполнением будет совет ассоциации.

Общий регламент по защите персональных данных, принятый Европейским союзом в 2016 году и вступивший в силу в 2018, — тема отдельной большой статьи. Да и про GDPR написано уже достаточно много,

ниже — ссылки на подробные материалы: [Rusbase](#): что нужно знать о новом регламенте; Блог [Digital Rights Center](#): GDPR — новые правила обработки персональных данных в Европе для международного IT-рынка; Оригинал регламента в удобном для чтения формате.

Я расскажу только про некоторые особенности GDPR и чем он отличается от №152-ФЗ. Экстерриториальное действие. Для российских компаний экстерриториальное действие закона о персональных данных не является чем-то принципиально новым. Но в отличие от отечественного закона GDPR напрямую предусматривает обязанность для компании, которая не находится в Европейском Союзе, назначить представителя на территории ЕС (статья 27 GDPR). Приятный бонус для государственных органов в Европе — возможность рассчитывать штраф для филиалов и представительств по доходу материнской компании.

Открытость и прозрачность. Цели, объемы и методы обработки персональных данных должны излагаться просто и доступно. По требованию европейских пользователей компании должны предоставлять полную информацию о том, какие данные обрабатываются, сколько хранятся и каким третьим лицам передаются. При этом, в отличие от России, уполномоченные в сфере защиты персональных данных органы государств ЕС активно штрафуют операторов персональных данных за «лишние» согласия, если у обработки данных есть иное основание. Такие согласия вводят пользователей в заблуждение, заставляют людей думать, что они могут отозвать согласие, и обработка прекратится.

Права субъекта данных. Граждане ЕС имеют право требовать от компании прекратить обработку их персональных данных, перенести в электронной форме данные из одного сервиса в другой, а также удалить личные данные по запросу во избежание их распространения, если это не противоречит общественным интересам.

Контроллер данных и процессор данных. Компании не всегда используют свои собственные ресурсы для хранения и обработки данных. Поэтому

большую юридическую ответственность несет тот, кто инициирует сбор, хранение и обработку данных (контроллер), а не компания-исполнитель (процессор).

По сравнению с №152-ФЗ GDPR дает гораздо больше прав субъектам персональных данных и в целом направлен на то, чтобы вернуть людям контроль над их цифровой жизнью и данными. Соответственно в рамках такого регулирования рынок данных пользователей не может существовать. Ниже к статье мы прикрепили материал по GDPR от PwC — GPS практика реализации требований. Судебные кейсы ВКонтакте vs Double Data.

В России публично о больших данных заговорили после подачи ВКонтакте иска против Double Data (ООО «Дабл»). В деле ранее также участвовало «Национальное бюро кредитных историй», но оно заключило с ВКонтакте мировое соглашение, в котором говорится, что «НБКИ обязуется изменить свои правоотношения с ООО «Дабл» таким образом, чтобы они не нарушали права истца, либо прекратить их».

Double Data использовала данные из открытых профилей пользователей ВКонтакте для анализа и продажи результатов такого анализа различным бюро кредитных историй и банкам. ВКонтакте под предлогом защиты пользователей подала иск, в котором просила суд признать базу пользовательских данных ВКонтакте интеллектуальной собственностью компании, а действия Double Data, соответственно, нарушением исключительного права ВКонтакте на базу данных. Double Data же считает, что данные пользователей не могут принадлежать ВКонтакте. Тем более пользователи сделали профили открытыми, а значит согласись, что любой посетитель страницы в соцсети может использовать данные этих профилей.

На сегодняшний день дело пересматривается после решения кассационной инстанции в Арбитражном суде Москвы, и его исход станет прецедентным и чрезвычайно важным для всего рынка больших данных России. Пока что суд приостановил разбирательство и назначил проведение

дополнительной экспертизы. Определение суда и список вопросов по делу можно прочитать [здесь](#).

Facebook vs Cambridge Analytica

Один из самых громких скандалов, связанных с пользовательскими данными, дело Facebook и Cambridge Analytica в Federal Trade Commission (Федеральная торговая комиссия США). Cambridge Analytica воспользовалась данными около 87 млн. пользователей Facebook для продвижения политической агитации, в том числе для президентской кампании Дональда Трампа. Около месяца назад, спустя 18 месяцев после скандала, FTC признала, что Cambridge Analytica обманывала пользователей Facebook и осуществляла сбор данных непрозрачно.

В итоге FTC наложила на Facebook штраф в \$5 млрд за нарушение правил хранения персональных данных. Также Facebook выплатила штраф ICO (Управление комиссара по информации Великобритании) в размере 500 000 фунтов стерлингов, поскольку в числе базы Cambridge Analytica также оказались пользователи Великобритании.

Однако на этом история не закончилась, и в конгрессе США всерьез озаботились проблемой того, как и для каких целей могут использоваться данные, которыми владеет Facebook. В частности, это касается вопросов дезинформации в политической рекламе.

Не менее значимым кейсом в области данных пользователей является иск к HiQ Labs от LinkedIn — и он гораздо больше напоминает разбирательство между российскими ВКонтакте и Double Data. HiQ Labs разрабатывает программы, которые помогают работодателям предсказывать поведение их сотрудников, в том числе планируемые увольнения. В качестве данных HiQ Labs использует открытые профили пользователей LinkedIn. Соцсеть заявляла, что в этом случае нарушаются права как пользователей сети, так и самой компании.

Апелляционный суд 9-го округа США посчитал, что данные профилей не могут принадлежать LinkedIn. И если пользователи делают свои профили

общедоступными, соответственно, данные этих профилей также доступны для всех. Получается, что HiQ Labs может продолжать использовать открытые данные в коммерческих целях, а единственный способ для LinkedIn ограничить в этом HiQ Labs — запретить создавать общедоступные профили, что, вероятно, нанесет ущерб самой соцсети.

Петиция на решение суда, поданная LinkedIn, была отклонена, что на данный момент лишний раз подтверждает легальность использования открытых данных пользователей в коммерческих целях. Тем не менее, разбирательство еще находится на предварительной стадии и нельзя однозначно утверждать, что решение будет окончательным.

Большие данные — технология обработки информации, которая превосходит сотни терабайт и со временем растет в геометрической прогрессии.

Такие данные настолько велики и сложны, что ни один из традиционных инструментов управления данными не может их хранить или эффективно обрабатывать. Проанализировать этот объем человек не способен. Для этого разработаны специальные алгоритмы, которые после анализа больших данных дают человеку понятные результаты.

В Big Data входят петабайты (1024 терабайта) или эксабайты (1024 петабайта) информации, из которых состоят миллиарды или триллионы записей миллионов людей и все из разных источников (Интернет, продажи, контакт-центр, социальные сети, мобильные устройства). Как правило, информация слабо структурирована и часто неполная, и недоступная.

Библиографический список:

1. Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 24.04.2020) "О персональных данных" // "Собрание законодательства РФ", 31.07.2006, N 31 (1 ч.), ст. 3451.

2. Федеральный закон от 27.07.2006 N 149-ФЗ (ред. от 08.06.2020) "Об информации, информационных технологиях и о защите информации" // "Собрание законодательства РФ", 31.07.2006, N 31 (1 ч.), ст. 3448.

3. Лескина Э.И. Применение блокчейн-технологий в сфере труда // Юрист. 2018. № 11. С. 25-30.

4. Лескина Э.И. Влияние развития нейронных сетей на трудовые отношения // Российская юстиция, 2020, № 8, С. 9-12.