

Правовое регулирование больших данных



Мария Абламейко,
доцент кафедры конституционного права Белорусского государственного университета, кандидат юридических наук, доцент

Использование в экономической, социальной, политической и других сферах жизнедеятельности больших данных (БД) приводит к необходимости введения правил и ограничений по работе с ними, в том числе неструктурированными, содержащими персональные сведения. Появлению БД предшествовало построение и развитие информационного общества, что закрепило за данными статус ценного экономического актива. Цифровая трансформация привела к экспоненциальному росту объемов генерируемой информации. Это связано в том числе с переходом от аналогового формата (бумажные документы, книги и др.) на цифровой и позволило аккумулировать гигантские массивы информации, кардинально изменило подходы к управлению ими во всех сферах общественных отношений. Развитие искусственного интеллекта и необходимость обучения нейронных сетей актуализировали потребность в колоссальных массивах данных, что обусловило их значимость. В связи с тем, что развитие технологий ИИ во многих развитых странах признано национальным приоритетом, все большее внимание уделяется как вопросам наличия данных, так и обеспечению легального доступа к ним.

На правовом уровне многие страны начинают принимать меры по созданию доступа к открытым данным, с одной стороны, и защите своих больших данных – с другой, так как их использование может прямо или опосредованно затрагивать аспекты обеспечения национальной и коллективной безопасности, включая безопасность личности. Международное сообщество стремится выработать общие подходы к регулированию big data таким образом, чтобы соблюсти баланс между стимулированием развития инноваций и безопасностью их функционирования. Решение данной задачи возможно с технической и юридической точки зрения. Внедрение технологий ИИ в общественные процессы не должно приводить к ущемлению прав и свобод человека независимо от экономического эффекта и коммерческого интереса.

На законодательном уровне следует решить ряд вопросов, включая определение терминологии, предоставление доступа к данным государственных структур (формирование института открытых данных), а также классификацию и разграничение big data от персональных данных методами обезличивания последних с целью надежной защиты прав и свобод человека в цифровом пространстве, прежде всего – неприкосновенности частной жизни.

Характеристика и соотношение понятий «информация» и «данные»

Термин «большие данные» чаще встречается в технической сфере и не имеет однозначного юридического определения. В юридической науке обсуждаются вопросы анализа правовых последствий применения больших данных. Особый интерес вызывают проблемы обеспечения конфиденциальности, защиты персональных данных, кибербезопасности, а также соблюдения норм антимонопольного законодательства крупными цифровыми платформами. Основной акцент исследователи делают на поиске ответов на технологические вызовы и выработке эффективных регуляторных мер.

Об этом писали В. Майер-Шенбергер и К. Кукьер в книге «Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим». Авторы подчеркивают необходимость наличия правовых рамок при обращении с данными, особенно в контексте приватности и предсказательной аналитики [1]. Л. Флориди рассматривает big data через призму этики с точки зрения морально обоснованных решений и прозрачности алгоритмов [2]. Риски и угрозы конфиденциальности, такие как слежка, раскрытие информации и дискриминация, анализирует американский юрист Р. Кало [3]. Деятельность платформ Google, Amazon и иных, использование ими данных с целью доминирования на рынке рассматривает А. Гавер. К вопросам правового определения обращались А. Минбалеев, Л. Терещенко и другие ученые.

Юридическое закрепление понятия «большие данные» как на международном, так и на национальном уровнях отсутствует. В силу этого отдельные аспекты регулируются дискретно – посредством норм, связанных с обработкой и защитой информации, персональных данных и цифровизацией.

Для выявления сущностной характеристики понятия следует обратиться к таким категориям, как «данные» и «информация». Последний термин широко распространен в нормативных правовых актах государств – участников СНГ, что обусловлено долгосрочной стратегией информатизации и цифрового развития в данном регионе.

При этом можно выделить несколько подходов. Например, в законодательных актах Российской Федерации, Кыргызской Республики, Республики Армения в сфере информации и информационных технологий термин раскрывается через категорию «данные», в нашей стране и Казахстане – через «сведения». В соответствии с Законом Республики Беларусь от 10.11.2008 г. №455-З «Об информации, информатизации и защите информации», информация – это сведения о лицах, предметах, фактах, событиях, явлениях и процессах, независимо от формы их представления.

В отличие от других стран, в Административном процедурно-процессуальном кодексе Республики Казахстан от 29.06.2020 г. №350-VI разъясняется понятие данных как информации в формализованном виде, пригодной для обработки.

Формулировка термина «большие данные» имеется в актах технического регулирования. В соответствии с ГОСТ Р ИСО/МЭК 20546-2021 «Национальный стандарт Российской Федерации. Информационные технологии. Большие данные. Обзор и словарь» (идентичен Международному стандарту ИСО/МЭК (ISO/IEC) 20546:2019 «Information technology – Big data – Overview and vocabulary», IDT), это большие массивы данных, отличающиеся главным образом такими характеристиками, как объем, разнообразие, скорость обработки и/или вариативность, которые требуют задействования технологии масштабирования для эффективного хранения, обработки, управления и анализа. Также отмечено, что термин «большие данные» широко применяется в различных значениях, например в качестве наименования технологии масштабирования, используемой для обработки больших массивов данных.

Четкое разграничение дефиниций «информация» и «данные» имеет принципиальное методологическое значение. С одной стороны, информация формируется на основе структурирования и интерпретации первичных данных. С другой – ее определение через категорию «сведения» требует симметричного введения понятия «данные» в правовое поле. По нашему мнению, сущность big data необходимо раскрывать именно через категорию данных, а не технологий, поскольку технологический инструментарий обработки таких массивов постоянно эволюционирует, в то время как сам объект регулирования остается неизменным.

Обобщенно можно сказать, что большие данные – это массивы структурированных и неструктурированных данных, обладающие определенными характеристиками, при обработке которых применяются различные технологии. Модель «3V: volume, velocity, variety», предложенная аналитиком компании Gartner

Д. Лэйни, легла в основу характеристики больших данных, в рамках которой рассматриваются такие параметры, как объем информации (ее количество, измеряющееся в гигабайтах, зеттабайтах или йоттабайтах); скорость (насколько быстро происходит генерирование и обработка); разнообразие (различные типы собранных из разных источников, что определяет формат таких данных (структурированные, неструктурированные и др.) [4].

Позднее усилиями экспертов, компаний и исследовательских сообществ в связи с развитием технологий и возрастающими потребностями рынка описанная трехкомпонентная модель пополнилась такими критериями, как veracity (достоверность) – качество данных, позволяющее повысить точность результатов, и value (ценность) – возможность определения скрытых закономерностей, прогнозирование, оптимизация процессов, персонализация предложений, то есть выгода, которую получает организация, что сформировало концепт 5V.

В настоящее время появляются новые характеристики, такие как volatility (изменчивость) – данные подвержены постоянной трансформации, что влияет на их однородность; visualization (визуализация) – делает big data доступными для человеческого восприятия; validity (валидность) – процесс проверки данных различных типов по критериям корректности и полезности для конкретного применения; vulnerability (уязвимость) и viscosity (сложность обработки данных) и др.

В связи с постоянным ростом объема данных, а также скорости обработки увеличивается их разнообразие. По структуре выделяются структурированные, полуструктурированные и неструктурированные. Данный критерий играет значительную роль в связи с наличием правового регулирования в сфере государственных информационных систем, а именно в части структурированных данных и отсутствия в части неструктурированных.

Правовой режим больших данных

Анализ законодательств стран СНГ позволяет выделить два основных режима, предполагающих полный и частичный доступ путем введения ограничительных мер для их использования: общедоступная информация и информация, доступ к которой ограничен. В условиях цифровизации правовой режим общедоступной информации трансформируется. Все страны находятся в поиске баланса между открытостью информации и обеспечением национальной безопасности. К общедоступной относится информация, доступ к которой не может быть ограничен на законодательном уровне, что гарантирует реализацию конституционного права граждан на доступ к информации, включая данные о деятельности госорганов и порядке предоставления услуг, в том числе электронных и др.

Более передовым можно считать опыт стран, где принят цифровой кодекс. Так, в Цифровом кодексе Кыргызской Республики от 31.07.2025 г. №178 установлен принцип открытости по умолчанию: «данные в общем пространстве цифровых данных являются открытыми данными». В Казахстане государство содействует развитию механизма открытости, достоверности, доступности, актуальности, полноты цифровых данных, их цифровой совместимости и свободного обмена цифровыми данными, что закреплено в Цифровом кодексе Республики Казахстан от 09.01.2026 г. №255-VIII.

Следует отметить, что большое количество данных собирается государственными органами в рамках выполнения возложенных на них полномочий, и их передача третьим лицам не предусмотрена без наличия на то законного основания. Существует множество инициатив по предоставлению открытого доступа к данным, выдвинутых национальными правительствами, межправительственными организациями, а также местными органами управления.

Международное сообщество призывает к открытию данных уже на протяжении длительного времени. В частности, в 2013 г. Департамент ООН по экономическим и социальным вопросам опубликовал «Руководящие принципы по открытым данным правительства» [5], а в 2014 г. была основана глобальная сеть «Открытые данные в интересах развития», которая преобразована в «Данные для развития сети» (Data for Development Network – D4D.net) для применения данных и цифровых технологий для решения важнейших задач развития Глобального Юга [6].

Многие государства уделяют приоритетное внимание разработке специализированных веб-платформ, открывающих публичный доступ к ведомственной информации. Подобные ресурсы успешно функционируют в Дании (Open Data DK), Ирландии (Ireland's Open Data Portal), Эстонии (Estonian Open data portal), а также в Саудовской Аравии, Австралии и др. [7].

Следует разграничивать понятия «большие» и «открытые» данные. Специфика big data

заключается в том, что они могут аккумулировать как общедоступные сведения, предоставляемые в публичном пространстве, так и информацию, оборот которой строго ограничен законодательством, включая персональные данные и иные конфиденциальные ресурсы. Информация с ограниченным доступом – это сведения, распространение и предоставление которых может повлечь за собой причинение вреда правам и законным интересам лиц, а также общественной, государственной или национальной безопасности в целом. К этой категории относятся охраняемые законом тайны (врачебная, банковская, коммерческая), а также персональные данные граждан. На уровне большинства юрисдикций сегодня сформирована развитая нормативная база, регламентирующая оборот персональных сведений субъектов. Особый акцент делается на обеспечении конфиденциальности в разнородных отраслях, оперирующих чувствительными массивами информации. При этом наиболее жесткий правовой режим традиционно присущ государственным секретам и иным сведениям, представляющим стратегическую ценность, раскрытие которых способно нанести ущерб национальной безопасности. В целом, аналогичные критерии градации и правовой защиты применимы и к феномену больших данных.

Технологии обработки больших массивов структурированных данных достаточно развиты как с технической точки зрения (в формате информационных ресурсов и систем), так и с юридической, что нашло отражение в действующих нормативных правовых актах. Основная коллизия заключается в том, что неструктурированные сведения крайне сложно отнести к какому-либо конкретному правовому режиму информации. Для легального и безопасного использования таких массивов целесообразно применять процедуру обезличивания, что позволит минимизировать или полностью исключить риски нарушения законодательства о персональных данных и конфиденциальности.

Таким образом, большие данные следует квалифицировать не только как колоссальные информационные массивы, но и как самостоятельную технологию их генерации, хранения и аналитической обработки. В связи с тем, что данный инструментарий относится к категории сквозных цифровых решений, его правовое регулирование неизбежно должно носить комплексный и межотраслевой характер. Регулирование технологии больших данных, как и технологий искусственного интеллекта, возможно в рамках определенных отраслей права в зависимости от сферы применения.

С точки зрения правового регулирования big data можно определить три подхода. В рамках первого следует фокусироваться на принятии специализированного акта. Второй подход предполагает внесение изменений в действующее законодательство в сфере информации. Третий – предусматривает дополнение законодательства о персональных данных, так эта сфера является наиболее критичной.

На наш взгляд, к этому вопросу следует подходить комплексно, используя все вышеуказанные стратегии. Определение терминологии, основных характеристик, правовых режимов в зависимости от сферы применения, правила предоставления доступа к большим данным государственных структур, а также процедуры обезличивания следует закрепить в Законе «Об информации, информатизации и защите информации». Вместе с тем отдельную главу, касающуюся вопроса анонимизации персональных данных, следует внести в Закон «О защите персональных данных». Впоследствии возможно формирование отдельного законодательного акта.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / В. Майер-Шенбергер, К. Кукьер; пер. с англ. И. Гайдюк. – М., 2014.
2. Floridi L. What is Data Ethics? / L. Floridi, M. Taddeo // Philosophical Transactions of The Royal Society a Mathematical Physical and Engineering Sciences. 2016. Vol. 374, №2083. Art. 20160360 // https://doi.org/10.1098/rsta.2016.0360.
3. Calo R. Consumer Subject Review Boards: A Thought Experiment / Big data and privacy making ends meet // https://fpf.org/wp-content/uploads/2013/08/Wu-Big-Data-Threats1.pdf.
4. Laney D. 3D Data Management: Controlling Data Volume, Velocity and Variety / Application Delivery Strategies, Meta Group, 2001.
5. Организация Объединенных Наций // https://publicadministration.desa.un.org/ru/topics/digital-government/ogd.
6. Data for Development Network // https://www.d4d.net.
7. Глобальный атлас регулирования искусственного интеллекта / под ред. А.В. Незнамова. – М., 2023.