ШАРОВА Д.Е.,
ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия, e-mail: d.sharova@npcmr.ru

МИХАЙЛОВА А.А.,
ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия, e-mail: a.mikhailova@npcmr.ru

ГУСЕВ А.В.,
к.т.н., ФГБУ «ЦНИИОИЗ» Минздрава России, Москва, Россия, e-mail: agusev@webiomed.ru

ГАРБУК С.В.,
к.т.н., НИУ ВШЭ, Москва, Россия, e-mail: garbuk@list.ru

ВЛАДЗИМИРСКИЙ А.В.,
д.м.н., ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия, ФГАОУ ВО Первый МГМУ им. И.М. Сеченова, Москва, Россия, e-mail: a.vladzimirsky@npcmr.ru

ВАСИЛЬЕВ Ю.А.,
к.м.н, ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия, e-mail: y.vasilev@npcmr.ru

АНАЛИЗ МИРОВОГО ОПЫТА В РЕГУЛИРОВАНИИ ИСПОЛЬЗОВАНИЯ МЕДИЦИНСКИХ ДАННЫХ ДЛЯ ЦЕЛЕЙ СОЗДАНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

DOI: 10.25881/18110193_2022_4_28

Аннотация.

В работе рассматривается мировой опыт регулирования использования медицинских данных для целей создания систем искусственного интеллекта (СИИ) с помощью методов машинного обучения. Для успешного внедрения СИИ в медицинскую практику и повышения эффективности принятия клинических и управленческих решений необходимы качественные наборы медицинских данных, для формирования которых в свою очередь требуется соответствующая нормативно-правовая база, учитывающая интересы всех участников на каждом из этапов разработки и использования СИИ.

Обзор зарубежных законодательств проводился для стран лидеров макрорегионов, которые были выбраны исходя из метрик рынка ИИ. На сегодняшний день существуют разные подходы к защите медицинских данных. Из них можно выделить отраслевой подход (США) и межотраслевой (ЕС). Для обеспечения надлежащего баланса между безопасностью пациента и возможностью сбора медицинских данных для разработчиков, необходимо формирование нормативно правовой базы как для межотраслевого, так и отраслевого регулирования.

Ключевые слова: *искусственный интеллект; машинное обучение; регулирование данных; законы о конфиденциальности; базы данных; медицинские данные; анонимизация.*

Для цитирования: *Шарова Д.Е., Михайлова А.А., Гусев А.В., Гарбук С.В., Владзимирский А.В., Васильев Ю.А. Анализ мирового опыта в регулировании использования медицинских данных для целей создания систем искусственного интеллекта на основе машинного обучения. Врач и информационные технологии. 2022; 4: 28-39. doi: 10.25881/18110193_2022_4_28.*

SHAROVA D.E.,
Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia, e-mail: d.sharova@npcmr.ru

MIKHAILOVA A.A.,
Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia, e-mail: a.mikhailova@npcmr.ru

GUSEV A.V.,
PhD, FRIHOI, Moscow, Russia, e-mail: agusev@webiomed.ru

GARBUK S.V.,
PhD, HSE University, Moscow, Russia, e-mail: garbuk@list.ru

VLADZYMYRSKYY A.V.,
DSc, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Sechenov First Moscow State Medical University, Moscow, Russia, e-mail: a.vladzimirsky@npcmr.ru

VASILEV Y.A.,
DSc, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia, e-mail: y.vasilev@npcmr.ru

AN ANALYSIS OF GLOBAL EXPERIENCE IN REGULATIONS ON THE USE OF MEDICAL DATA FOR ARTIFICIAL INTELLIGENCE SYSTEMS DEVELOPMENT BASED ON MACHINE LEARNING

DOI: 10.25881/18110193_2022_4_28

Abstract.

The paper covers international experience in regulating the use of medical data for the development of artificial intelligence systems (AI) using machine learning methods. High-quality medical data sets are required for successful implementation of AI in medical practice and for higher efficiency of clinical and managerial decision-making. Such data sets are impossible to acquire, store and use without appropriate legal and regulatory framework that takes into account the interests of all participants at each stage of the development and use of AI.

The review of foreign legislations was carried out for the countries — leaders of the macro-regions, which were selected based on the higher metrics of the AI market. Today, there are different approaches to protecting medical data, with the most well-known being industry and cross-industry approaches (USA and EU respectively). In order to keep a proper balance between patient safety and the possibility of collecting medical data for developers, a regulatory framework for both cross-being industry and cross-industry regulation needs to be formed.

Keywords: *artificial intelligence; machine learning; data regulations; privacy laws; databases; healthcare data; anonymization.*

For citation: *Sharova D.E., Mikhailova A.A., Gusev A.V., Garbuk S.V., Vladzymyrskyy A.V., Vasilev Y.A. An analysis of global experience in regulations on the use of medical data for artificial intelligence systems development based on machine learning. Medical doctor and information technology. 2022; 4: 28-39. doi: 10.25881/18110193_2022_4_28.*

ВВЕДЕНИЕ

Внедрение систем искусственного интеллекта (СИИ) является одним из ключевых трендов цифровой трансформации здравоохранения [1]. Пандемия COVID-19 способствовала существенному росту интереса к использованию искусственного интеллекта (ИИ) в здравоохранении, стимулировала изменения на законодательном уровне в большинстве стран и увеличение инвестиций, а также привлекла внимание общественности [2]. По данным Reports and Data, размер рынка ИИ-систем для медицины и здравоохранения в 2021 году достиг 7 млрд долларов. Ожидается, что среднегодовой прирост составит 46,7%, т.е. 215,53 млрд долларов в 2030 году [3].

Считается, что применение СИИ поможет лучше анализировать медицинскую информацию, в т.ч. неструктурированные медицинские записи, оценивать изменение данных пациента во времени, прогнозировать риск развития заболеваний, выявлять аномалии в данных и т.д. За счет этого можно улучшить эффективность принятия клинических и управленческих решений. Таким образом, СИИ обладают теоретическим потенциалом для преобразования многих аспектов ухода за пациентами и повышения качества медицинских услуг, включая сокращение расходов и снижение нагрузки на медицинских работников.

Создание СИИ на современном этапе часто подразумевает применение алгоритмов машинного обучения (МО), которые в свою очередь требуют соответствующие наборы данных (НД) [4]. В мире наблюдается повсеместный рост накапливаемых медицинских данных, включая электронные медицинские карты (ЭМК), уровень использования которых увеличился с 9% в 2008 г. до 44% в 2012 г. и продолжает расти [3]. Одним из наиболее рациональных подходов к формированию НД является создание условий для повторного использования уже накопленных сведений в различных медицинских информационных системах. Такой подход требует соответствующего нормативного регулирования сбора и использования ранее накопленных данных разработчиками СИИ и другими заинтересованными лицами.

Основной сложностью такого регулирования является поиск баланса между удобством доступа к обезличенным данным для целей исследований и разработок в сфере ИИ и обеспечением должного уровня конфиденциальности и безопасности этих данных [5]. На рисунке 1 изображена схема развития систем ИИ с точки зрения регулирования медицинских баз данных.

Необходимо отметить, что законодательное регулирование и соответственно доступ к медицинским данным неравномерны по миру. Они напрямую зависят от государственной политики в части оборота обезличенных медицинских данных в каждой отдельно взятой стране. На рисунке 2 отображено распределение наборов данных между странами — первоисточниками [6].

На данный момент ключевой барьер для внедрения СИИ в здравоохранении заключается не столько в технологической готовности к решению той или иной задачи обработки данных, сколько в возможности легитимного использования создаваемых технологий в повседневной клинической практике [1]. Для широкого внедрения они должны быть одобрены регулирующими органами на национальных уровнях, интегрированы в систему здравоохранения и стандартизованы [7]. Это относится и к регулированию использования медицинских данных, без которых функционирование таких систем не представляется возможным. На это предположительно уйдет гораздо больше времени, чем потребуется для развития самих технологий [1].

Таким образом, для развития СИИ на основе методов МО важно обеспечить контролируемую доступность качественных наборов медицинских данных, что позволит создавать новые алгоритмы и программные продукты на основе ИИ, повышая тем самым ценность и востребованность их в реальной клинической практике [8].

Целью настоящего исследования явился анализ международного опыта в нормативном регулировании вопросов, связанных с использованием медицинских данных для исследований и разработок на основе методов МО.

МАТЕРИАЛЫ И МЕТОДЫ

Для основных зарубежных макрорегионов (Европа, США, Латинская Америка и Азия) был проведен обзор законодательства, относящегося к использованию данных в сфере ИИ, и выявлены основные нормативные механизмы обеспечения компромисса между удобством предоставления доступа к данным и обеспечением их конфиденциальности.

Рисунок 1 — Развитие ИИ с точки зрения безопасности и доступности медицинских данных.

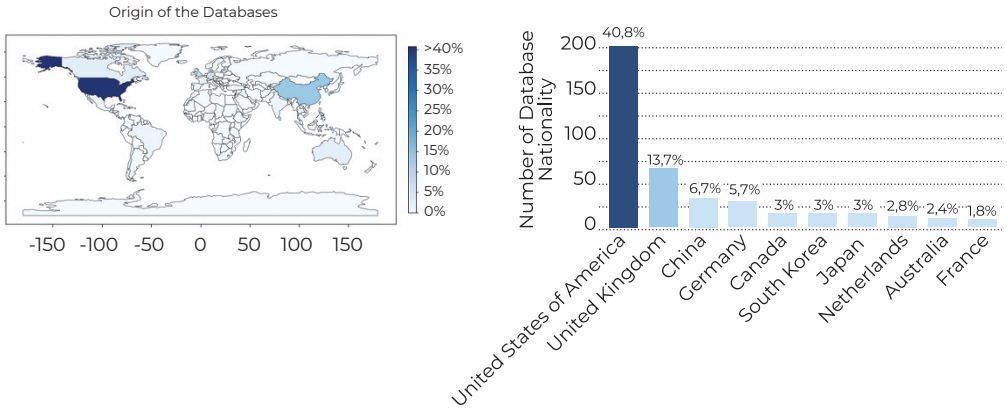

Рисунок 2 [6] — Страны — первоисточники баз данных.

Кроме того, был выполнен анализ существующих показателей рынка ИИ в здравоохранении для этих макрорегионов. Проанализированы ключевые метрики рынка ИИ [9]: суммы привлеченных инвестиций в ИИ-компании; количество созданных ИИ компаний в странах, представляющих наибольший интерес относительно их размера, влияния, или других факторов. Анализ был выполнен на период с 2016 по 2020 гг.

ОБЗОР ЗАКОНОДАТЕЛЬСТВА РАЗЛИЧНЫХ СТРАН

В таблице 1 приведены результаты анализа рынка по лидерам в рассмотренных регионах. Только в трех странах (США, Китай и Индия) объем частных инвестиций превысил 9 млрд долларов в год. Самое большое число новых компаний с 2016 по 2020 год возникло в США и ЕС.

Как правило, системы национального нормативного регулирования строятся на ряде правил, главная цель которых состоит в обеспечении вывода на рынок безопасных и эффективных изделий, включая лекарственные средства, вакцины и устройства медицинского назначения [10]. Особенностью СИИ является стремление регуляторов найти баланс между стимулированием инвестиций для вывода на рынок новых перспективных продуктов и обеспечением безопасности и конфиденциальности данных, необходимых для МО. Сохранение баланса с целью улучшения медицинского обслуживания между вторичным использованием данных других пациентов и

Таблица 1 — Показатели рынка по лидерам макрорегионов

Регион	Лидер	Количество новых компаний 2016–2020	Частные инвестиции на 2020 год, млрд долларов
-	США	4772	23,6
Европа	Германия	2074 (во всех странах ЕС)	2,1 (во всех странах ЕС)
Латинская Америка	Бразилия	178	0,03
Восточная Азия	Китай	841	9,9
Юго-Восточная Азия	Сингапур	300	0,3
Южная Азия	Индия	924	9,4

конфиденциальностью персональной информации вызывает ряд проблем. Например, на уровне отдельных пациентов проблемой является понимание, в какой степени их персональные данные подвергаются вторичному использованию и какие элементы этих данных задействованы; кто может получить доступ к данным; насколько анонимность данных эффективна и полна; могут ли эти данные потенциально использоваться для нанесения вреда пациентам; можно ли изменить их данные; используются ли их данные для финансовой выгоды других лиц и повлияет ли изменение политики конфиденциальности данных в ближайшем или отдалённом будущем на получаемую ими помощь [11]. На уровне медицинских учреждений, между которыми происходит обмен медицинскими данными, встаёт вопрос «осознания ответственности» за владение персональными данными пациентов, в общем случае препятствующий такому обмену. На государственном уровне ответственность за медицинские ошибки в условиях широкого использования систем ИИ, в основу которых положены медицинские данные большого количества пациентов, становится ещё более неоднозначной [12].

Далее будут рассмотрены особенности регулирования персональных медицинских данных в основных мировых макрорегионах.

1. США

США является лидером по количеству инвестиций в ИИ и количеству баз медицинских данных в мире. Правовая база США построена на федеральной системе, каждый штат имеет свой собственный свод законов, правил и положений, касающихся вопросов защиты персональных данных. Основной федеральный закон, применяемый к медицинским данным — это HIPAA или Health Insurance Portability and Accountability Act [13]. Он определяет, как должна быть защищена личная информация, относящаяся к сфере здравоохранения и медицинского страхования.

В США отсутствует общеотраслевое (единое) законодательство о защите данных [14]. Вместо этого выпускаются отдельные отраслевые законодательные акты о защите информации для здравоохранения, образования, связи или финансовых услуг. При этом федеральные законы носят высокоуровневый характер, допуская дополнительное точечное регулирование на уровне отдельных штатов.

Для здравоохранения в 1996 г. принят «Акт о передаче и защите данных учреждений здравоохранения» (Health Insurance Portability and Accountability Act, HIPAA). Данный закон содержит положения, предусматривающие защиту и обеспечение конфиденциальности закрытой медицинской информации (PHI). Определение PHI охватывает широкий спектр информации, включая данные страховки и информацию об оплате, информацию о диагнозе, клиническом обслуживании и результатах обследований, например снимках и анализах. Правила HIPAA применяются к учреждениям, подпадающим под действие закона: больницам, поставщикам медицинских услуг, корпоративным организациям здравоохранения, научно-исследовательским учреждениям и страховым компаниям, которые работают непосредственно с пациентами и их данными. Требование HIPAA о защите PHI также распространяется на деловых партнеров этих учреждений.

Общий регламент защиты персональных данных ЕС (англ. *General Data Protection Regulation, GDPR*) является значительно более всеобъемлющим и горизонтально применяется ко всем секторам экономики, ко всем видам персональных данных и всем, кто контролирует или обрабатывает эти данные. Согласно источникам HIPAA

обеспечивает значительно более слабую защиту персональных данных чем GDPR, несмотря на то что меры защиты являются самыми сильными относительно других отраслей в США [15].

HIPAA запрещает поставщикам медицинских услуг и предприятиям здравоохранения раскрывать защищенную информацию кому-либо, кроме пациента и уполномоченных представителей пациента без их согласия. Однако существуют исключения, например, при наличии соглашения между медицинскими организациями и их деловыми партнерами (англ. *business associate*) [16], которые оказывают услуги для медицинской организации. Эти партнеры могут получить доступ к личной медицинской информации и, если обозначено в договоре, могут обезличить/деидентифицировать (англ. *de-identify*) данные в соответствии со стандартами HIPAA и коммерциализировать их [17]. Договорное разрешение на обезличивание данных может быть частью сделки и влиять на стоимость услуг, оказываемых данной организацией в рамках договора. Продажи таких данных не отслеживаются, о них не нужно сообщать и они, вероятно, происходят повсеместно [18]. Это помогает индустрии США активно развивать свои ИИ системы, однако вызывает беспокойство насчет конфиденциальности личных данных пациентов [17]. На рисунке 3 представлена схема возможных вариантов коммерциализации медицинских данных в США.

Отметим, что в HIPAA четко сформулированы требования по обезличиванию и по возможностям реидентификации. Существует два метода для достижения деидентификации в соответствии с HIPAA: метод “экспертного определения”; или удаление идентификаторов физического лица или родственников, работодателей или членов семьи физического лица [19]. Данные, обезличенные согласно этим техникам, не всегда будут считаться таковыми в рамках GDPR.

HIPAA применяется только к данным, хранящимся у традиционных поставщиков медицинских услуг. Множество компаний, не связанных напрямую со здравоохранением, рассматривают медицинские данные как имеющие большую ценность. Некоторые компании поощряют потребителей самостоятельно собирать и обрабатывать данные из мобильных медицинских приложений и носимых устройств, не обеспечивая достаточную прозрачность для субъектов этих данных [20]. Существует мнение, что узкие секторальные меры защиты личных данных не являются достаточной мерой и вызывают все большее беспокойство [21]. При этом следует учитывать, что секторальный подход позволяет лучше подстраиваться под отдельную сферу и использует более конкретные формулировки, а универсальный помогает защищать трансграничные данные о здоровье, например, если данные имеют медицинский характер, но получены не в медицинском учреждении.

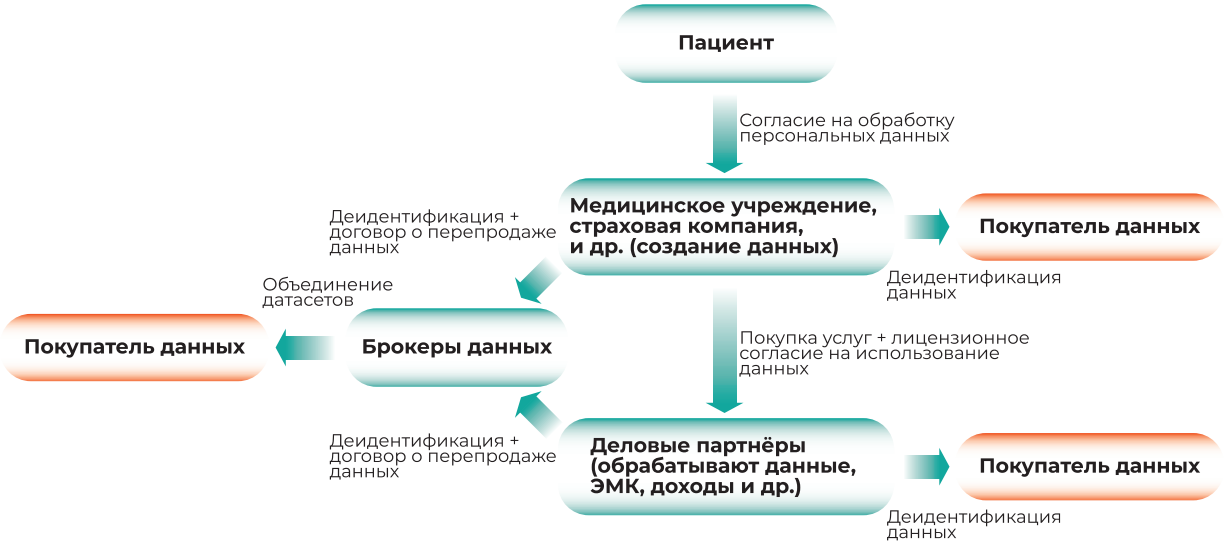

Рисунок 3 — Схема коммерциализации медицинских данных в США.

2. Европейский союз

Наборы данных из Европы занимают относительно небольшую часть рынка, а частные инвестиции на 2020 год примерно в 5 раз ниже, чем в Индии и Китае, и в 10 раз ниже, чем в США [22]. Возможные причины такого различия в инвестициях могут быть обусловлены: сложностью координации нормативно правовых актов в отношении данных между 27 государствами ЕС; их фрагментарным применением на национальном/субнациональном уровне; разнообразием приоритетов, проблем и подходов среди членов ЕС; сложностью удовлетворения установленных требований относительно использования персональных данных, в частности — медицинских данных.

Основным документом для регулирования персональных данных в ЕС является Общий регламент защиты персональных данных или GDPR [23], вступивший в силу 25 мая 2018. Его основными задачами являются унификация и защита персональных данных граждан стран — членов ЕС. GDPR значительно усиливает индивидуальный контроль субъектов персональных данных и поднимает мировую планку в этом отношении. Данный регламент стимулировал создание нового законодательства о регулировании данных по всему миру. Некоторые из стран (Аргентина, Бразилия и другие) приняли законы похожие на GDPR по различным причинам, например, облегченный режим торговли — использование личных данных резидентов Европы возможно только если ЕС счел регулирование данных в отдельно взятой стране адекватным (Япония, Канада, и другие) [24].

Однако именно этот документ стал причиной обеспокоенности по поводу его влияния на европейскую конкурентоспособность и опасности формирования избыточных ограничений на стыке технологий и общества как основного барьера для цифровых инноваций. Далее мы рассмотрим некоторые аспекты GDPR, которые препятствуют формированию НД, их использование в научных целях, и какие из ключевых понятий нуждаются в формализации.

Сложность практического применения требований GDPR во многом обуславливается недостаточно конкретным характером формулировок и требований. За некоторыми исключениями GDPR запрещает любую обработку неанонимизированных персональных данных без согласия субъекта данных [25]. Полностью анонимизированные (англ. *anonymized*) или обезличенные данные больше не считаются персональными данными, если риск их разглашения минимален. Сложность определения личных и неличных данных состоит в том, что концептуальные границы, разделяющие их согласно GDPR, сильно размыты. Кроме того, эти данные динамичны, а однажды анонимизированные данные могут снова стать личными, если появятся новые технологии, позволяющие повторно идентифицировать их [26; 27]. Исследования показывают, что даже обезличенные наборы данных зачастую не удовлетворяют требованиям, установленным в GDPR [28].

В GDPR вводится понятие псевдонимизация (англ. *pseudonymized*) [29] — мера защиты конфиденциальности, которая оставляет возможной идентификацию субъекта с использованием каких-либо косвенных мер. В соответствии с GDPR даже при удалении идентифицирующих полей данные по-прежнему считаются в большинстве случаев персональными данными, и это не исключит другие необходимые меры защиты данных. Медицинские данные попадают под специальную категорию персональных данных (англ. *sensitive data*), поэтому в случае коммерческого использования требуют согласия [26].

Проблема повторного использования медицинских данных усложняется ещё и потому, что в соответствии с GDPR данные должны «собираться для конкретных, отчетливых, законных целей и не обрабатываться в последующем несовместимым с этими целями образом» [30]. Таким образом, если при сборе данных изначально не преследовалась цель разработки алгоритмов ИИ, то использование этих НД для решения задач обучения интеллектуальных медицинских систем может оказаться проблематичным. Следующий спорный момент возникает в преамбуле №75 где указано, что «риск для прав и свобод физических лиц разной степени вероятности и серьезности может возникать..., когда обработка охватывает большое количество персональных данных и затрагивает большое количество субъектов данных». При этом не уточняется, какое количество персональных данных и субъектов данных можно считать «большим».

Существуют исключения, когда согласие субъекта данных для категории специальных персональных данных не требуется, например, если «обработка [данных] необходима для целей

архивирования в общественных интересах, научных или исторических исследований или статистических целей» [31]. Это благоприятствует исследованиям на территории ЕС, однако не является исключением для трансграничной передачи данных. В таком случае Европейская комиссия оценивает то, какие страны обеспечивают адекватный уровень защиты персональных данных. Список этих стран весьма ограничен, и большинство стран, с которыми ведутся совместные исследования, в него не входят. Это сильно усложняет процесс совместных исследований и требует описывать те меры, которые использовались для псевдонимизации данных [32].

В целом, к понятиям, требующим дальнейшей формализации, следует отнести:

- показатели уровня и критерии достаточности анонимизации и пседонимизации;
- значения рисков, наступающих при нарушении конфиденциальности, целостности и доступности персональных данных;
- объективные критерии возрастания уровня конфиденциальности данных по мере их накопления и обобщения.

Наиболее остро проблема формализации встает для понятий «анонимизация» и «псевдонимизация». В частности, не определена модель угроз, в которой были бы заданы возможности потенциального злоумышленника по восстановлению персональной принадлежности предварительно обезличенных данных. Без фиксации таких возможностей в общем случае невозможно с уверенностью судить о качестве выполненного обезличивания и предоставлять гарантии невозможности обратного восстановления данных.

Вопрос использования обезличенных данных и четкости в их регулировании в рамках GDPR остается открытым, до сих пор не существует единого подхода в этой области, на что, например, обращали внимание в Германии [33]. Это оказывает негативное влияние на Европейскую экономику в области обучения систем ИИ. Несмотря на то, что окончательная версия GDPR стала более детализированной относительно анонимизации и информированного согласия, она все равно подвержена критике, как со стороны индустрии, так и со стороны исследователей [34; 35].

Преодолению этих и других сложностей имплементации GDPR в значительной мере способствует разработка нормативно-технических документов (международных и национальных стандартов), уточняющих и конкретизирующих отдельные требования и понятия.

3. Азия

Далее рассмотрены три региона Азии в силу их культурных и экономических различий: Юго-Восточная Азия, Восточная Азия и Южная Азия.

3.1. Восточная Азия

Лидерство в разработке новейших технологий стало центральным элементом усиливающегося соперничества между Китаем и США. Позиция китайского правительства заключается в том, что технологии ИИ имеют решающее значение для экономической и национальной безопасности Китая [36]. Частные инвестиции в ИИ в Китае на 2020 год более чем в два раза ниже, чем в США (9,5 млрд долларов и 23,6 млрд долларов, соответственно). Важно отметить, что Китай имеет значительные государственные инвестиции в ИИ. При этом как центральные, так и местные органы власти в Китае тратят значительные средства на НИОКР [37].

Правительство Китая активно продвигает использование больших данных в медицине, которые считаются стратегическим национальным ресурсом. В 2016 году Государственный совет Китая опубликовал официальное уведомление о разработке и использовании больших данных в секторе здравоохранения [38]. Согласно документу, это поможет улучшить здравоохранение в Китае. В нем были определены цели развития, задачи и организационная структура [39].

В 2021 году вступил в силу новый закон Китая о защите личной информации (англ. *Personal Information Protection Law of the People's Republic of China, PIPL*) [40]. Он стал первым всеобъемлющим законодательным актом Китая, регулирующим защиту персональной информации физических лиц. PIPL во многом схож с GDPR. Однако он содержит и некоторые новшества, которые еще предстоит интерпретировать в будущем с учетом сложившейся практики применения этого закона и понять, как это отразится на формировании и использовании баз медицинских данных.

PIPL содержит положения, которые требуют получения согласия субъекта персональных данных для последующей их обработки. Одно

из исключений, когда согласие не требуется, — “другие обстоятельства, предусмотренные законами и административными правилами” [40]. Это может дать китайскому правительству больше гибкости в интерпретации и применении этого закона в будущем. Согласно мнению некоторых юристов и экспертов по кибербезопасности PIPL, в отличие от GDPR, больше связан с национальными интересами и национальной безопасностью [41].

Аналогично GDPR обезличенные данные не считаются персональными данными согласно PIPL. Однако даже они могут рассматриваться как «важные данные» или «большие медицинские данные». Это приводит к более строгому контролю за хранением и передачей таких данных [42].

Некоторые страны Восточной Азии также пересматривают свою законодательную базу для одобрения ЕС в рамках GDPR [43]. Южная Корея, например, пересмотрела свое законодательство, чтобы иметь возможность подать заявку на одобрение ЕС. На данный момент две страны получили авторизацию от ЕС — Южная Корея и Япония.

3.2. Юго-Восточная Азия

Лидером этого региона является Сингапур, с частными инвестициями в $315 млн на 2020 год при размере населения в 5,7 млн человек. Сингапур позиционирует себя как место для глобального предпринимательства в области инноваций. Прямые инвестиции и партнерские отношения с игроками глобального венчурного капитала сыграли решающую роль в создании уникальной экосистемы, которая направлена на оптимизацию шансов стартапов на региональный успех [44].

3.3. Южная Азия

В южной Азии однозначным лидером является Индия, занимающая третье место по частным инвестициям в мире на 2016–2020 гг. Системы ИИ в медицине могут решить такие серьезные проблемы, как нехватка медицинского персонала на душу населения и доступ к медицинской помощи [45]. Однако существует острая нехватка объемных и качественных баз медицинских данных, что значительно тормозит процесс развития в этой области [45].

Законодательная база по защите личных данных в Индии находится на этапе активных изменений, так как сейчас она фрагментарна. Аналогично ЕС Индия формирует Personal Data Protection Bill (PDPB), который станет первым национальным законом, посвящённым сугубо защите персональных данных [46].

4. Латинская Америка

Экосистема ИИ и законодательная база по контролю личных данных в Латинской Америке находится на стадии активного формирования. Мировые лидеры ИИ налаживают свои исследовательские связи с регионом, растет число компаний. В Бразилии количество новых ИИ компаний в 2016–2020 гг. составило 178, что в несколько раз превышает этот показатель для таких стран как Россия, Индонезия и другие [22].

Бразилия, Мексика, Чили и Аргентина разработали или сейчас разрабатывают официальные национальные стратегии в области ИИ и совершенствуют законодательную базу. Защита данных в Латинской Америке активно развивается и по причине того, что во многих странах действует законодательство, которое не было реформировано в соответствии с требованиями ЕС. Это ограничивает возможности взаимодействия указанных макрорегионов. На данный момент две страны — Уругвай и Аргентина, соответствуют требованиям ЕС согласно GDPR [24].

Общий регламент защиты персональных данных (англ. *General Personal Data Protection Law, LGPD*) Бразилии вступил в силу в 2020 году [47]. Этот закон устанавливает и защищает права и свободы относительно персональных данных физических лиц. Многие положения LGPD совпадают с GDPR, но имеются и различия. Например, в GDPR использование данных для исследований является исключением, тогда как в LGPD это является правовым основанием к использованию с желательным обезличиванием, если оно возможно [48; 49].

ОБСУЖДЕНИЕ

Законодательство, регулирующее оборот обезличенных медицинских данных в разных странах, сильно варьируется и находится на разных этапах своего развития. На данный момент большинство баз медицинских данных доступно в США (40,8%) и Китае (13,7%) [6]. Благодаря в том числе проактивному законодательному регулированию в этих странах, удалось создать

соответствующие национальные экосистемы сбора и хранения данных в сфере здравоохранения, к которым ученые и разработчики ИИ-систем могут получить доступ, что облегчает обучение алгоритмов методами МО. По мнению авторов указанной статьи, именно наличие таких данных стало одним из решающих факторов, обеспечивших США и Китаю лидерство на глобальном рынке ИИ-систем для здравоохранения.

Система регулирования США с секторальным подходом ориентирована в большей степени на интересы разработчиков систем ИИ, но уступает в защите личных данных. В ЕС, напротив, защита личных данных является приоритетом. Таким образом, если рассматривать авторизацию ЕС в качестве индикатора качества защиты персональных данных, то страны, получившие эту авторизацию, можно считать относительно развитыми с точки зрения защиты личных данных. При секторальном подходе могут существовать более четкие требования по обращению с данными, что значительно упрощает такие процессы, как, например, обезличивание. Сбалансированный подход реализуют страны Азии, включая Японию (10-ое место по частным инвестициям в мире за 2015–2020 года) и Южную Корею (13-ое место). Вероятно, именно эти страны лучше других смогли найти оптимум защиты медицинских данных (подход ЕС), интересов индустрии (подход США) и интересов государства (подход Китая).

Анализ существующего мирового опыта в регулировании медицинских данных с точки зрения возможности формировать обезличенные наборы данных для целей МО и развития искусственного интеллекта показывает, что в мире нет единого подхода к данному вопросу.

Вместе с этим, проведенный анализ подтверждает следующую тенденцию: усложнение законодательства в части сбора и обработки обезличенных данных препятствует развитию достижений в сфере искусственного интеллекта.

Напротив, ослабление регулирования в области защиты персональных данных и либерализация требований к сбору обезличенных данных является одним из существенных факторов, обеспечивающих ускоренное развитие технологий ИИ для здравоохранения с достижением лидирующих позиций на этом рынке.

ЗАКЛЮЧЕНИЕ

Обеспечение надлежащего доступа к медицинских данным с соблюдением конфиденциальности и безопасности является сложной и актуальной задачей. Выполненный нами анализ позволил сформулировать следующие выводы:

- Межотраслевое и отраслевое регулирования по отдельности имеют свои плюсы и минусы. Вероятно, только при сочетании двух подходов возможно сбалансировать интересы пациентов, государства и индустрии.
- Если ЕС сочтет регулирование данных в стране адекватным, это может послужить хорошим стимулом для развития данной области и будет способствовать привлечению дополнительных инвестиций.
- Необходима формализация таких неотъемлемых понятий, характеризующих НД для систем ИИ, как анонимизация, псевдонимизация и др. Это непростая задача, однако ее решение может помочь индустрии и исследователям лучше понимать требования и быстрее выполнять многие из этапов подготовки медицинских данных к использованию.

Развитие в Российской Федерации национальной нормативной правовой и нормативно-технической базы, обеспечивающей сбор обезличенных медицинских данных и формирование на их основе качественных НД с регламентированным порядком доступа для российских научно-исследовательских организаций и компаний-разработчиков ИИ-систем, будет способствовать совершенствованию отечественных интеллектуальных медицинских технологий и развитию отрасли здравоохранения в целом.

ЛИТЕРАТУРА/REFERENCES

1. Davenport T., Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. Royal College of Physicians. 2019; 6(2): 94-98. doi: 10.7861/futurehosp.6-2-94.
2. Leslie D, et al. Does “AI” stand for augmenting inequality in the era of covid-19 healthcare? BMJ. 2021; 372. doi: 10.1136/bmj.n304.
3. Artificial Intelligence in Healthcare Market Size & Share 2030. Reports and data. 2022. https://www.reportsanddata.com/report-detail/artificial-intelligence-in-healthcare-market.

4. Павлов Н.А. и др. Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике. // Digital Diagnostics. — 2021. — Т.2. — №1. — С.49-66. [Pavlov NA, et al. Etalonnye medicinskie datasety (MosMedData) dlya nezavisimoj vneshnej ocenki algoritmov na osnove iskusstvennogo intellekta v diagnostike. Digital Diagnostics. 2021; 2(1): 49-66. (In Russ.)] doi: 10.17816/DD60635.
5. Winter J.S. AI in healthcare: data governance challenges. J Hosp Manag Health Policy. 2021; 5. doi: 10.21037/jhmhp-2020-ai-05.
6. Celi LA, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities — A global review. PLOS Digital Health. 2022; 1(3). doi: 10.1371/journal.pdig.0000022.
7. Зинченко В.В. и др. Стандартизация в области регулирования технологий искусственного интеллекта в российском здравоохранении // Казанский медицинский журнал. — 2021. — Т.102. — №6. — С.923-933. [Zinchenko VV, et al. Standartizaciya v oblasti regulirovaniya tekhnologij iskusstvennogo intellekta v rossijskom zdravoohranenii. Kazanskij medicinskij zhurnal. 2021; 102(6): 923-933. (In Russ.)] doi: 10.17816/KMJ2021-923.
8. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. The Lancet. 2020; 395: 1579-1586. doi: 10.1016/S0140-6736(20)30226-9.
9. Zhang D, et al. The AI Index 2021 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University. Stanford, 2021.
10. Ethics and Governance of Artificial Intelligence for Health: WHO guidance. World Health Organization. 2021.
11. Jaremko JL, et al. Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. Can Assoc Radiol J. 2019; 70(2): 107-118. doi: 10.1016/j.carj.2019.03.001.
12. Шарова Д.Е. и др. К вопросу об этических аспектах внедрения систем искусственного интеллекта в здравоохранении // Digital Diagnostics. — 2021. — Т.2. — №3. — С.356-368. [SHarova DE, et al. K voprosu ob eticheskih aspektah vnedreniya sistem iskusstvennogo intellekta v zdravoohranenii. Digital Diagnostics. 2021; 2(3): 356-368. (In Russ.)] doi: 10.17816/DD77446.
13. Health Insurance Portability and Accountability Act of 1996. Public law. 1996. https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996.
14. Data Protection Laws and Regulations Report 2022 USA. The International Comparative Legal Guides. 2022. https://iclg.com/practice-areas/data-protection-laws-and-regulations/usa.
15. Data Privacy and Protection Relating to Healthcare in Europe, the United States and Brazil. Latin Lawyer. 2020. https://www.lexology.com/library/detail.aspx?g=99b83b76-3f2f-4b23-a5c3-30ad576af369.
16. Covered Entities and Business Associates. U.S. Department of Health & Human Services. https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html.
17. McGraw D, Petersen C. From Commercialization to Accountability: Responsible Health Data Collection, Use, and Disclosure for the 21st Century. Appl Clin Inform. 2020; 11(2): 366-373. doi: 10.1055/ s-0040-1710392.
18. Tanner A. Our Bodies, Our Data: How Companies Make Billions Selling Our Medical Records. Beacon Press, 2017.
19. Methods for De-identification of PHI. U.S. Department of Health & Human Services. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.
20. Grundy Q, et al. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. The BMJ. BMJ Publishing Group. 2019; 364. doi: 10.1136/bmj.l920.
21. Terry N. Existential challenges for healthcare data protection in the United States. Ethics Med Public Health. 2017; 3(1): 19-27. doi: 10.1016/j.jemep.2017.02.007.
22. Zhang D, et al. The AI Index 2022 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University. Stanford, 2022.
23. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. 2016. https://gdpr-info.eu.
24. Adequacy decisions. European Commission. https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en.
25. Art. 6 GDPR — Lawfulness of processing — General Data Protection Regulation (GDPR). Official Journal of the European Union. 2016. https://gdpr-info.eu/art-6-gdpr/.
26. Durovic M, Montanaro M. Data Protection and Data Commerce: Friends or Foes? European Review of Contract Law. 2021; 17(1): 1-36. doi: 10.1515/ercl-2021-000.

27. Wrobel M. Anonymized data — curse or blessing of data protection?! Taylor Wessing LLP. 2020. https://www.lexology.com/library/detail.aspx?g=1517d319-4184-4d49-b3a9-d0e99da65019.
28. Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun. 2019; 10(1). doi: 10.1038/s41467-019-10933-3.
29. Art. 4 GDPR — Definitions — General Data Protection Regulation (GDPR). Official Journal of the European Union. 2016. https://gdpr-info.eu/art-4-gdpr/.
30. Art. 5 GDPR — Principles relating to processing of personal data — General Data Protection Regulation (GDPR). Official Journal of the European Union. 2016. https://gdpr-info.eu/art-5-gdpr/.
31. Art. 9 GDPR — Processing of special categories of personal data — General Data Protection Regulation (GDPR). Official Journal of the European Union. 2016. https://gdpr-info.eu/art-9-gdpr/.
32. Войниканис Е.А. Большие (персональные) данные: проблема баланса интересов // Журнал Суда по интеллектуальным правам. — 2021. — Т.34. — №4. — С.19-27. [Vojnikanis EA. Bol'shie (personal'nye) dannye: problema balansa interesov. ZHurnal Suda po intellektual'nym pravam. 2021; 34(4): 19–27. (In Russ.)]
33. BfDI nutzt erstmals Konsultationsverfahren // der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit. 2020. https://www.bfdi.bund.de/SharedDocs/Pressemitteilungen/DE/2020/03_Konsultationsverfahren.
34. van Veen EB. Observational health research in Europe: understanding the General Data Protection Regulation and underlying debate. Eur J Cancer. 2018; 104: 70-80. https://doi.org/10.1016/j.ejca.2018.09.032.
35. Peloquin D, DiMaio M, Bierer B, Barnes M. Disruptive and avoidable: GDPR challenges to secondary research uses of data. European Journal of Human Genetics. 2020; 28(6): 697-705. https://doi.org/10.1038/s41431-020-0596-x.
36. Allen CG. Understanding China's AI Strategy. Center for a New American Security. 2019. https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.
37. Colvin TJ, Liu I, Babou TF, Wong GJ. A Brief Examination of Chinese Government Expenditures on Artificial Intelligence R&D. Institute for Defense Analyses. 2020. https://www.ida.org/research-and-publications/publications/all/a/ab/a-brief-examination-of-chinese-government-expenditures-on-artificial-intelligence-r-and-d.
38. China to boost big data application in health and medical sectors. The State Council of the People's Republic of China. 2016. http://english.www.gov.cn/policies/latest_releases/2016/06/24/content_281475379018156.htm.
39. Zhang L, et al. Big data and medical research in China. BMJ. 2018; 360. doi: 10.1136/bmj.j5910.
40. Translation: Personal Information Protection Law of the People's Republic of China — Effective Nov. 2021; 1. Stanford University. https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021.
41. Zhu J. The Personal Information Protection Law: China's Version of the GDPR? Columbia Journal of Transnational Law. 2022. https://www.jtl.columbia.edu/bulletin-blog/the-personal-information-protection-law-chinas-version-of-the-gdpr.
42. Chen D, Wang K. At a glance: data protection and management of health data in China. Ropes & Gray LTD. 2022. https://www.lexology.com/library/detail.aspx?g=fd2bb402-33d5-4ba7-85a7-c5383cb11526.
43. The Asia Pacific Privacy Guide 2020-2021 // Deloitte Asia Pacific Limited. 2020. https://www2.deloitte.com/id/en/pages/risk/articles/ap-privacy-guide-2020-2021.html.
44. How Singapore brings together the best in innovation and investment to drive start-up growth. Investment Monitor and Singapore Economic Development Board. 2022. https://www.investmentmonitor.ai/tech/how-singapore-brings-together-the-best-in-innovation-and-investment-to-drive-start-up-growth.
45. Parry C. M., Aneja U. AI in Healthcare in India: Applications, Challenges and Risks. Chatham House, International Affairs Think Tank. 2020. https://www.chathamhouse.org/2020/07/artificial-intelligence-healthcare-insights-india-0/3-ai-healthcare-india-applications.
46. Data Protected India. Talwar Thakore & Associates. 2022. [https://www.linklaters.com/en/insights/data-protected/data-protected-india.
47. LGPD Brazil — General Personal Data Protection Act. Data Protection National Authority. 2018. https://lgpd-brazil.info.
48. Article 7: Chances of Carrying Out Personal Data Processing — Chapter 2 — Processing of Personal Data — LGPD Brazil. Data Protection National Authority. 2018. https://lgpd-brazil.info/chapter_02/article_07.
49. Article 11: Processing of Sensitive Personal Data — Chapter 2 — Processing of Personal Data — LGPD Brazil. Data Protection National Authority. 2018. https://lgpd-brazil.info/chapter_02/article_11.