

А.В. ГУСЕВ^{1,2,3}, канд. тех. наук, директор по развитию бизнеса, agusev@webiomed.ai
ORCID: 0000-0002-7380-8460

Е.М. АСТАПЕНКО⁴, канд. тех. наук, директор департамента регулирования обращения лекарственных средств и медицинских изделий, AstapenkoEM@minzdrav.gov.ru

И.В. ИВАНОВ⁵, д-р. мед. наук, генеральный директор, ivi1976@bk.ru ORCID: 0000-0003-0971-853X

Т.В. ЗАРУБИНА⁶, д-р. мед. наук, профессор, директор института цифровой трансформации медицины, t_zarubina@mail.ru ORCID: 000-0003-3989-4884.

Б.А. КОБРИНСКИЙ⁷, д-р мед. наук, профессор, заведующий отделом систем интеллектуальной поддержки принятия решений, bak@isa.ru ORCID: 0000-0002-3459-8851

Принципы формирования доверия к системам искусственного интеллекта для сферы здравоохранения

¹ ООО «К-Скай», 185031, Российская Федерация, Республика Карелия, г. Петрозаводск, наб. Варкауса, д. 17. Limited liability company "K-Sky", 17, Varkaus embankment, Petrozavodsk, Republic of Karelia, 185031, Russian Federation.

² ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, 127254, Российская Федерация, г. Москва, ул. Добролюбова, д. 11. The Federal State Budget Institution "The Central Research Institute for Health Organization and Informatics" of the Ministry of Health of the Russian Federation, 11, Dobrolyubova St., Moscow, 127254, Russian Federation.

³ ГБУЗ г. Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы», 127051, Российская Федерация, г. Москва, ул. Петровка, д. 24, стр. 1. State Budgetary Institution healthcare "Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Department of Healthcare of the City of Moscow", 24, bld 1, Petrovka St., Moscow, 127051, Russian Federation.

⁴ Министерство здравоохранения Российской Федерации, 127994, Российская Федерация, г. Москва, ГСП 4, Рахмановский пер., д. 3. Ministry of Health of the Russian Federation, 3, Rakhmanovsky pereulok St., City Service Post Office-4 Moscow 127994 Russian Federation.

⁵ ФГБУ «Всероссийский научно-исследовательский и испытательный институт медицинской техники» Росздравнадзора, 115478, Российская Федерация, г. Москва, Каширское шоссе, д. 24, стр. 16. Federal State Budget Institution "Russian Scientific and Research Institute for Medical Engineering" of Roszdravnadzor, 24, bld. 16, Kashirskoye highway, Moscow, 115478, Russian Federation.

⁶ ФГАОУ ВО «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации, 117997, Российская Федерация, г. Москва, ул. Островитянова, д. 1. Federal State Autonomous Educational Institution of Higher Education "N.I. Pirogov Russian National Research Medical University" of the Ministry of Health of the Russian Federation, 1, Ostrovityanova St., Moscow, 117997, Russian Federation.

⁷ Федеральный исследовательский центр «Информатика и управление» РАН, 119333, Российская Федерация, г. Москва, ул. Вавилова, д. 44, корп. 2. Federal Research Center "Informatics and Management" of the Russian Academy of Sciences, 44/2, Vavilova St., Moscow, 119333, Russian Federation.

Ключевые слова: цифровое здравоохранение; искусственный интеллект; доверенный искусственный интеллект; машинное обучение; системы поддержки принятия решений; требования к интеллектуальным медицинским системам

Для цитирования: Гусев А.В., Астапенко Е.М., Иванов И.В., Зарубина Т.В., Кобринский Б.А. Принципы формирования доверия к системам искусственного интеллекта для сферы здравоохранения // Вестник Росздравнадзора. – 2022. – № 2. – С. 25–33.

For citation: Gusev A.V., Astapenko E.M., Ivanov I.V., Zarubina T.V., Kobrinskii B.A. Principles for building confidence in artificial intelligence systems for healthcare // Vestnik Roszdravnadzora. – 2022. – Vol. 2. – P. 25–33.

Gusev A.V., Astapenko E.M., Ivanov I.V., Zarubina T.V., Kobrinskii B.A. Principles for building confidence in artificial intelligence systems for healthcare

The introduction of artificial intelligence technologies is a key area for the digital transformation of healthcare. It is believed that the use of AI-based software will revolutionize healthcare by improving clinical decision-making, digital image analysis, risk management, predicting and monitoring the correctness of prescribed treatment. At the same time, many publications state that there is currently a serious problem of insufficient data on the clinical efficacy and safety of AI products brought to the market. Manufacturers make claims about product efficacy, but few provide credible evidence. In this regard, over the past few years, regulators, supervisory authorities, experts and professional communities have intensified the discussion around the need for thoughtful and strict regulation of issues of responsible attitude to the development and application of AI technologies in healthcare and clinical medicine. Based on the analysis of various documents and recommendations, we defined 4 key principles for a responsible attitude to the development, application and supervision of the use of AI technologies in Russian healthcare: "Confidence in the manufacturer", "Data confidence", "Model confidence" and "Product confidence". What is more, we proposed evaluation criteria and suggestions for the development of legal and technical regulation.

Keywords: digital health; artificial intelligence; confidence in artificial intelligence; machine learning; decision support systems; requirements for intelligent medical systems

Внедрение технологий искусственного интеллекта (ИИ) является одним из ключевых направлений цифровой трансформации здравоохранения. Считается, что применение программного обеспечения на основе ИИ произведет революцию в здравоохранении, в т.ч. благодаря улучшению принятия клинических решений, анализу цифровых изображений, управлению рисками, прогнозированию и контролю правильности назначаемого лечения. Вместе с этим существует множество публикаций, констатирующих, что в настоящее время имеется серьезная проблема недостаточности данных о клинической эффективности и безопасности выводимых на рынок ИИ-продуктов. Многие из них содержат многообещающие заявления производителей об эффективности, однако единицы предоставляют действительно заслуживающие доверия доказательства клинической эффективности и безопасности. В этой связи за последние несколько лет в мире со стороны регуляторов, надзорных органов, экспертов и профессиональных сообществ активизировалась дискуссия о необходимости более продуманного и строгого регулирования вопросов ответственного отношения к разработке и применению технологий ИИ в сфере здравоохранения и клинической медицины. Основываясь на анализе разнотипных документов и рекомендаций, мы сформулировали четыре ключевых принципа ответственного отношения к разработке, применению и надзору за использованием технологий ИИ в российском здравоохранении: «Доверие к производителю», «Доверие к данным», «Доверие к моделям» и «Доверие к продукту», а также предложили критерии оценки и сформулировали предложения по развитию нормативного и технического регулирования.

Введение

В настоящее время одной из лидирующих технологий в развитии цифрового здравоохранения является искусственный интеллект (ИИ). Это понятие предполагает построение систем на основе знаний предметной области, полученных от экспертов и из литературных источников или извлеченных из данных. Созданные на этой основе системы поддержки принятия решений (СППР) могут быть ориентированы на решение клинических и управленческих задач. Считается, что внедрение различных продуктов, использующих ИИ, произведет революцию в здравоохранении, в т.ч. благодаря консультативной помощи при принятии клинических решений, анализе цифровых изображений, управлении рисками, прогнозировании и контроле безопасности назначаемого лечения [1].

Системы на основе методов искусственного интеллекта позволяют решать задачи, которые традиционно считаются «творческими», они могут включать то, что принято называть рассуждениями, реализуемыми в интеллектуальных системах на основе аргументации. Большие успехи были достигнуты в использовании систем искусственного интеллекта для решения диагностических задач [2, 3]. В последнее время значительный прогресс наблюдается в распознавании медицинских изображений

с использованием нейросетевого подхода глубокого обучения, получившего название компьютерного зрения [4, 5].

Основной предпосылкой интереса к технологиям ИИ является рост заболеваемости, в особенности хроническими неинфекционными заболеваниями. Вызванный этим рост затрат на здравоохранение создает необходимость в более эффективном отраслевом управлении, в т.ч. благодаря автоматизации рутинных функций. Также имеются технологические предпосылки: рост вычислительной мощности компьютеров, развитие облачных вычислений, быстрое накопление больших данных. Это послужило основой для развития технологии машинного обучения в различных задачах лечебно-диагностического процесса.

Проведенное в 2020 г. исследование Accenture показало, что в США и Европе уже 69% медицинских организаций начали пилотировать или внедрять технологии ИИ [6]. Согласно опросу 500 руководителей здравоохранения, который провела Optum в 2021 г., 96% заявили, что ИИ будет играть важную роль в их усилиях по сохранению здоровья, причем 85% подтвердили, что в их организации есть стратегия ИИ, а почти половина (48%) руководителей заявили, что уже внедрили различные ИИ-инструменты в своих организациях

и планируют развивать их дальше. Основными направлениями для применения ИИ являются: автоматизация доклинических и административных процессов (72% голосов руководителей), внедрение продуктов виртуального ухода за пациентами (41%), диагностика и прогнозная аналитика (40%), анализ медицинских изображений (36%). Вместе с этим 94% руководителей считают тему ответственного отношения и доверия к ИИ продуктам одной из самых важных [7]. Аналогичный опрос, проведенный Optum годом ранее, также выявил озабоченность медицинского профессионального сообщества безопасностью и оправданностью использования ИИ, в особенности при принятии клинических решений. Так, 73% опрошенных специалистов указали на проблему отсутствия прозрачности того, как ИИ принимает решение и еще 69% заявили, что озабочены ролью человека в процессе принятия решений [8].

Применение технологий ИИ для улучшения различных аспектов клинической практики, получившее в зарубежной литературе также название дополненной медицины (augmented medicine), предполагает конвергенцию человеческого и искусственного интеллекта [9, 10, 11]. Однако, отсутствие правовой базы, определяющей понятие ответственности в случае принятия или отклонения рекомендаций интеллектуальной системы, делает врача уязвимым перед потенциальными юридическими последствиями при использовании ИИ [12]. В ИИ для медицины существует ряд «горячих точек» [13], среди которых и доверие к системам ИИ.

Понятие доверенного или надежного (англ. trustworthy) искусственного интеллекта было первоначально зафиксировано в европейском «Руководстве по этике для доверенного (надежного) ИИ» (Ethics guidelines for trustworthy AI, 2019 [14]). В нем указывается, что доверенный ИИ состоит из трех компонентов, которые должны соблюдаться на протяжении всего жизненного цикла системы: (1) соответствие законам

и нормативным документам, (2) согласие с этическими принципами и (3) надежность как с технической, так и с социальной точки зрения. В проекте стандарта ISO/IEC TR 24 028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence [14] более детально рассматриваются подходы, связанные с надежностью или установлением доверия к системам ИИ через прозрачность, объяснимость, управляемость и т. д.; технические ловушки и типичные связанные с ними угрозы и риски для систем ИИ; возможные приемы и методы смягчения негативных последствий на разных этапах жизненного цикла систем; подходы к оценке и достижению доступности, надежности, точности, безопасности, защищенности и конфиденциальности систем ИИ.

Применение технологий ИИ для улучшения различных аспектов клинической практики предполагает конвергенцию человеческого и искусственного интеллекта.

Таким образом, доверенный ИИ должен включать: объяснимость и качество решений, надежность исходной информации (база знаний, набор данных), безопасность пациента, конфиденциальность и информационную безопасность, психологическую совместимость пользователя с СППР.

Вопросы безопасности в условиях применения систем ИИ

Нельзя отрицать определенный положительный опыт применения интеллектуальных СППР в медицине. Однако следует отметить наличие многочисленных публикаций, констатирующих, что в настоящее время в сфере цифрового здравоохранения в целом, и применения технологий ИИ в частности, имеется серьезная проблема

недостаточности данных о клинической эффективности и безопасности разрабатываемых моделей машинного обучения и выводимых на рынок на их основе программных продуктов. В mHealth ряд продуктов содержат многообещающие заявления производителей об эффективности, однако единицы из них предоставляют действительно заслуживающие доверия доказательства клинической эффективности и безопасности [16, 17]. Например, Larsen M.E. и соавт. приводят результаты исследования 1435 мобильных приложений для психического здоровья, из которых 73 были включены в окончательный анализ, который выявил, что 64% приложений заявляли о своей эффективности в диагностике состояния психического здоровья, однако только два из них предоставили доказательства клинической эффективности [17].

В то же время интеллектуальный анализ текста со словами «компьютер», «система», «рабочая станция» и «сеть» использовался для изучения дескрипторов инцидентов с произвольным текстом для выявления инцидентов, связанных со сбоями в работе медицинских ИТ. Guy Martin и соавт. приводят результаты ретроспективного анализа 10-летних данных об инцидентах в Англии и Уэльсе, связанных с применением программного обеспечения для медицинских целей. Авторы выявили 2627 инцидентов, связанных со сбоями ИТ в области здравоохранения. Из них 82% не причинили вреда пациентам, 13% причинили незначительный вред, 4% причинили вред средней степени тяжести, 1% причинили серьезный вред и менее 1% способствовал смерти больного. При этом 75% инцидентов были предотвратимыми. Авторы пришли к выводу, что ошибки в ПО несомненно являются потенциальным источником вреда, вследствие чего необходимо рассматривать ИТ в области здравоохранения с точки зрения фундаментальных принципов безопасности пациентов, разрабатывать более эффективные методы выявления последствий ошибок и принимать меры для снижения их вероятности и риска [18].

L. Wynants и соавт. представили результаты анализа литературы относительно достоверности и полезности моделей прогнозирования развития COVID-19, в котором авторы сделали вывод, что почти все изученные прогностические модели плохо описаны и содержат высокий риск систематических ошибок, а заявленные разработчиками метрики их точности, вероятнее всего, слишком оптимистичны [19].

К аналогичным выводам пришли M. Roberts и соавт., которые опубликовали результаты анализа 62 исследований, посвященных разработке моделей машинного обучения для анализа рентгенограмм грудной клетки и изображений компьютерной томографии (КТ) грудной клетки для выявления признаков COVID-19, в котором авторы сделали вывод, что ни одна из моделей не пригодна для потенциального применения в реальной клинической практике из-за методологических недостатков и предубеждений [20].

A. Wong и соавт. опубликованы результаты независимого исследования точности работы модели на основе ИИ для прогнозирования сепсиса. Данная модель входит в число 20 запатентованных ИИ-алгоритмов, предназначенных для выявления и прогнозирования различных заболеваний, созданных компанией Epic Systems, крупнейшего американского поставщика одноименной системы ведения электронных медицинских карт (ЭМК). Исследование было проведено в условиях реальной клинической практики Мичиганского университета, где были изучены 38 тыс. историй болезни. Авторы пришли к выводу, что в реальности чувствительность модели составила 33%, точность 12%. Из 6 971 предупреждений системы о наличии сепсиса только 843 были правильными [21].

Подходы к формированию доверия к системам искусственного интеллекта

В последние несколько лет в мире со стороны регуляторов, надзорных органов, экспертов и профессиональных

Таблица 1. Перечень руководящих документов, рассматривающих вопросы ответственного отношения к системам искусственного интеллекта для здравоохранения

Название документа, год выпуска и организация-автор	Краткое описание
Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) – Discussion Paper and Request for Feedback [27]. Весна 2019, FDA США	Первый документ FDA по теме особого отношения к регулированию ИИ в здравоохранении, выпущенный в формате «дискуссионного документа». В нем FDA предложило рынку риск-ориентированный подход, согласно которому ПО на базе технологий ИИ воспринимается как особый объект для регулирования, имеющий перспективы и в то же время потенциальные опасности по сравнению с обычными программными медицинскими изделиями. FDA предложила компаниям-разработчикам ИИ-продуктов включать в свои заявки на регистрацию медицинских изделий планы предполагаемых модификаций и использовать два специальных документа «Предварительную спецификацию SaMD» (SPS) и «Протокол изменения алгоритмов» (ACP). SPS должен определять виды изменений, которые планируется внести, в то время как ACP подробно описывает процедуры, которых следует придерживаться для внесения этих изменений
Four Principles of Explainable Artificial Intelligence [28]. Август 2020, Национальный институт стандартов и технологий США	Предлагает четыре принципа, на которых должны быть основаны ИИ-системы: 1. Объяснение (Explanation). Системы ИИ должны предоставлять доказательства или причины, на основании которых были приняты те или иные решения. 2. Целенаправленность (Meaningful). Системы ИИ должны представлять объяснения для выходных данных, понятные (адаптированные) для разных групп пользователей. 3. Точность объяснения (Explanation Accuracy). Объяснение должно надежно (правильно) отражать суть процессов, производимых системой ИИ для генерирования результатов. 4. Пределы знаний (Knowledge Limits). Система работает только в условиях, для которых она была разработана, или при достижении надлежащей достоверности в выходных данных
Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan, [29]. Январь 2021, FDA США	В этом документе FDA представила рекомендации разработчикам ИИ-систем по разработке документа «Протокол изменений алгоритма», с помощью которого регулятор может оценить прозрачность разработки ИИ-алгоритмов с целью одобрения выпуска на рынок решения, которое получит право итеративно обновляться, сохраняя при этом способность обеспечивать безопасность для пациентов и не требуя вместе с этим постоянной регистрации новых версий
The Use of Artificial Intelligence in Health Care: Trustworthiness (ANSI/CTA-2090), [30]. Февраль 2021, Американская ассоциация потребительских технологий (CTA)	Предлагает три базовых принципа доверия к ИИ: 1. Человеческое доверие . Включает в себя быстроту реагирования разработчика на запросы пользователя продукта, пользовательский опыт, степень автономности ИИ-решения и уровень сложности его использования. 2. Доверие к информации и ее защищенности . Для создания ИИ-продукта должны быть использованы качественные, верифицированные и полные датасеты, он должен отвечать требованиям конфиденциальности и обеспечивать безопасность данных. 3. Доверие со стороны регулирующих органов . Продукт должен соответствовать требованиям нормативных документов, разработанных регулируемыми органами
Horizon Scanning Assessment Report – Artificial Intelligence [31]. Август 2021, Международная коалиция регуляторных органов в области лекарственных средств (International Coalition of Medicines Regulatory Authorities, ICMRA)	Рекомендует регуляторам следующее: 1. Воспринимать ИИ-систему, предназначенную для применения в клинической медицине, как программное медицинское изделие (SaMD) высокого класса риска. 2. Из-за сложности и новизны тем, связанных с ИИ, контролирующим органом настоятельно рекомендуется поддерживать ИИ-разработчиков при их обращениях за советом, в т.ч. по вопросам юридического, технического и этического аспектов разработки и эксплуатации ИИ. 3. Клиническая валидация ИИ требует прозрачности и объяснимости работы ИИ-алгоритмов, что в свою очередь дает основания требовать доступ регулятора к описанию работы алгоритма и набором данных, которые были использованы для машинного обучения. 4. В случае обновления ПО или ИИ-алгоритмов требуется повторное тестирование и другие исследования, чтобы убедиться в воспроизводимости работы ИИ. 5. В случае, если при работе ИИ в клинической медицине будут выявлены новые данные о пользе/риске лекарственного средства, такая информация должна быть предоставлена в надзорные органы. 6. Тестирование правильности работы ИИ – зона ответственности разработчика. В идеале он должен принимать меры для усиления управления разработкой, непрерывного управления данными и обеспечивать меры безопасности и конфиденциальности
Ethics and governance of artificial intelligence for health [32]. Июнь 2021, B03	Предлагает шесть принципов регулирования ИИ в здравоохранении: 1. Защита автономии человека : люди должны сохранять контроль над системами здравоохранения и медицинскими решениями. Неприкосновенность частной жизни и конфиденциальность должны быть защищены. Пациенты должны быть проинформированы о применении ИИ-системы через согласие. 2. Содействие благополучию и безопасности людей и общественным интересам . ИИ-системы должны соответствовать нормативным требованиям безопасности, точности и эффективности для определенных вариантов использования. 3. Обеспечение прозрачности, объяснимости и понятности . Прозрачность требует, чтобы была опубликована достаточная информация об особенностях ИИ-продукта. Такая информация должна быть легкодоступной. 4. Воспитание ответственности . Хотя технологии ИИ выполняют определенные задачи, заинтересованные стороны несут ответственность за то, чтобы ИИ использовался в соответствующих условиях и был обучен должным образом. 5. Обеспечение инклюзивности и справедливости . Инклюзивность требует, чтобы ИИ для здоровья был разработан независимо от возраста, пола, этнической принадлежности, особенностей трудовой деятельности или других характеристик, защищенных кодексами прав человека. 6. Продвижение адаптивного и устойчивого ИИ . Дизайнеры, разработчики и пользователи должны постоянно и прозрачно оценивать приложения ИИ во время практического применения, чтобы определить, адекватно ли и правильно ли ИИ реагирует на ожидания и требования. Регуляторные органы и компании должны реагировать на ожидаемые сбои на рабочем месте, включая обучение медицинских работников адаптации к использованию систем ИИ

Таблица 1. Окончание

Название документа, год выпуска и организация-автор	Краткое описание
Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, [33]. Июнь 2021, Счетная палата США (U.S. Government Accountability Office – GAO)	Предлагает четыре принципа ответственного использования ИИ: 1. Управление. Должны быть четко определены цели и задачи внедрения ИИ, роли и обязанности, ценности. К работе следует привлекать персонал с навыками работы с системами ИИ разного профиля. Должно быть разработано управление рисками и технические спецификации систем ИИ. ИИ-проект должен быть прозрачен для внешней экспертизы. 2. Данные. Источники должны быть задокументированы, надежность и полнота заполнения данных оценена, переменные тщательно выбраны. Должна быть изучена зависимость / независимость данных. Обеспечены меры безопасности и конфиденциальности. 3. Производительность. Должны быть задокументированы все модели, используемые в ИИ-системе, и метрики точности и производительности, выходные данные. 4. Мониторинг. Должен быть внедрен и задокументирован непрерывный мониторинг работы системы ИИ, определены допустимые пределы отклонений. Должна быть оценена полезность работы ИИ-системы, а также критерии масштабирования
Good Machine Learning Practice for Medical Device Development: Guiding Principles [34] Октябрь 2021, FDA США	Предлагает 10 руководящих принципов, призванных содействовать безопасной и эффективной разработке медицинских устройств и систем, использующих технологии машинного обучения: 1. Многоаспектная экспертиза применяется в процессе всего жизненного цикла программного продукта, что обеспечивает глубокое понимание интеграции модели в клинический процесс, а также преимущества и связанные с ними риски для пациентов, результатом чего является обеспечение безопасности и эффективности медицинских устройств. 2. Проектирование модели реализовано с учетом обеспечения безопасности. 3. Наборы данных должны представлять репрезентативную выборку пациентов. 4. Наборы данных, используемые при обучении модели и для ее тестирования, независимы. 5. Отобранные эталонные наборы данных основаны на лучших доступных практиках. 6. Проектирование модели адаптировано к имеющимся данным и отражает предполагаемое использование устройства/системы. 7. Основное внимание уделяется совместной работе команды «человек – искусственный интеллект». 8. Испытания демонстрируют работу системы в клинически значимых условиях. 9. Пользователям предоставляется четкая, необходимая информация. 10. Внедренные модели контролируются на предмет эффективности, а риски переобучения контролируются

сообществ активизировалась дискуссия о необходимости более продуманного и строгого регулирования вопросов ответственного отношения к разработке и применению технологий ИИ в сфере здравоохранения и клинической медицины. Основными проблемами в работе ИИ-систем названы повышенный риск причинения вреда здоровью пациентов в силу некачественных решений, сложность интерпретации решений при машинном обучении (проблема «черного ящика»), повышенные риски самообучающихся алгоритмов, которые способны изменять свою работу вследствие появления новых клинических данных, в том числе, получаемых во время эксплуатации, проблемы кибербезопасности, проблемы смещения данных, которые в свою очередь приводят к асимметрии между теми данными, на которых обучались ИИ-модели и данными реальной клинической практики [22, 23, 24, 25].

В итоге ряд авторов приходят к мнению, что с точки зрения нормативного

регулирования контроля качества и эффективности ИИ-систем к ним необходимо подходить более осторожно, по аналогии с обработкой биоматериалов человека [26].

Краткий перечень разработанных в мире к настоящему времени документов, рассматривающих вопросы ответственного отношения к ИИ в медицине и здравоохранении, представлен в *таблице 1*.

Базовые принципы ответственного отношения к применению искусственного интеллекта в здравоохранении Российской Федерации

Правовые основы внедрения и развития технологий ИИ в Российской Федерации определены Указом Президента РФ от 10.10.2019 № 490¹, которым была утверждена «Национальная стратегия развития искусственного интеллекта в Российской Федерации на период до 2030 года». Документ предусматривает ряд базовых принципов, среди которых

¹ Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации». URL: <http://kremlin.ru/acts/bank/44731>.

в области доверия к ИИ особое значение имеют следующие:

- ИИ-продукты должны быть безопасными**, в том числе иметь минимальный риск причинения вреда здоровью пациентам (раздел III, п.19 б).
- ИИ-продукты должны быть понятными с точки зрения принятия решений**. Необходимо обеспечить объяснение работы ИИ-продуктов, а также доступность информации о применяемых в них алгоритмах (раздел III, п.19 в).

Основываясь на анализе международных документов и рекомендаций, мы сформулировали четыре ключевых принципа ответственного отношения к разработке,

применению и надзору за использованием технологий ИИ в российском здравоохранении:

- Доверие к производителю**. Ключевым требованием является наличие у производителя системы менеджмента качества (СМК), сертифицированной доверенной независимой организацией-аудитором. Только продукты такого производителя целесообразно поддерживать для применения в реальной клинической практике. Важным критерием представляется наличие у производителя успешного опыта по разработке систем ИИ для медицинской предметной области.

Таблица 2. Принципы ответственного отношения к системам искусственного интеллекта для здравоохранения, требования и предлагаемые механизмы регулирования

Принципы ответственного отношения	Требования	Регламентирующие документы и надзорные мероприятия
1. Доверие к производителю	Высокий уровень компетенций ключевых сотрудников производителя СИИ	Свидетельства о регистрации Роспатента на программные продукты. Наличие публикаций в рецензируемых изданиях, иллюстрирующих эффективность разработок в области медицинских систем ИИ
	Компания-производитель должна внедрить систему менеджмента качества в соответствии со стандартом ISO 13485	Наличие сертификата соответствия СМК производителя на соответствие стандарту ISO 13485. Периодический независимый аудит СМК производителя
2. Доверие к данным	Производитель должен получать «сырую» информацию для последующего создания наборов данных строго в обезличенном виде и при условии письменного соглашения с операторами данных	Запрос надзорным органом реестра соглашений поставщиков «сырых» данных у производителя
	Производитель должен использовать надежные источники информации для создания качественных наборов данных для машинного обучения	Отраслевой стандарт с требованиями к качеству данных, используемых для создания моделей машинного обучения, тестирования и валидации моделей Репозиторий проверенных наборов данных
3. Доверие к моделям	Доступность отчетов о создании построенных производителем моделей для независимой оценки	Отраслевой стандарт с требованиями к отчету о валидации модели. Включение отчетов о машинном обучении в регистрационное досье на медицинское изделие. Публикация производителями отчетов в открытом доступе, либо публикация статей на основе отчетов в научной литературе
	Независимая валидация моделей	Добровольное проведение производителем независимой валидации моделей. Публикация результатов независимой валидации в виде отчета или статьи
4. Доверие к продукту	Регистрация программного продукта в качестве медицинского изделия должна сопровождаться гибкой схемой определения класса потенциального риска	Обязательное наличие у производителя регистрационного удостоверения на СИИ
	Проведение проспективных исследований эффективности и безопасности	Открытая публикация результатов исследований эффективности и безопасности СИИ в условиях реальной клинической практики
	Подключение применяемых в здравоохранении СИИ к системе государственного мониторинга	Подключения СИИ, имеющей регистрационное удостоверение, к федеральной системе государственного мониторинга

2. **Доверие к данным.** При машинном обучении должны быть использованы только качественные источники данных и сами наборы данных, соответствующие определенному уровню качества. Непрозрачность источников информации, использованных для создания наборов данных, способов и принципов их формирования, следует считать незаслуживающими доверия.
3. **Доверие к моделям.** Интеллектуальные системы, основанные на знаниях, должны характеризоваться точностью предлагаемых решений на уровне высококвалифицированного эксперта. Модели, полученные с использованием технологий машинного обучения, включенные в выпускаемые на рынок продукты, должны соответствовать определенным метрикам качества и точности работы построенных моделей.
4. **Доверие к продукту.** Система ИИ должна пройти регистрацию как медицинское изделие. Должно быть обеспечено доверие к каждой выпускаемой новой версии СИИ. Ситуация, при которой производитель самостоятельно, без внешнего доверенного надзора, выпускает новые версии СИИ на рынок, является настораживающей в плане неуточненного характера рисков работы продукта.

Сформированные нами критерии доверия и подходы к надзору за их исполнением представлены в *таблице 2*.

Важно обратить внимание, что сама по себе декларация принципов доверия мало что изменит. В этой связи необходимо не просто обсуждать и актуализировать эти принципы, но и предусмотреть механизмы их реализации. Одним из эффективных и целесообразных вариантов продвижения ИИ в широкую практику может стать создание реестра СИИ, который должен формироваться надзорным органом совместно с отраслевым регулятором. Такой реестр можно было бы рекомендовать органам управления здравоохранением, медицинским, страховым и фармацевтическим организациям для внедрения и использования продуктов в проектах цифровой трансформации здравоохранения.

Заключение

По нашему мнению, создание доверенных систем ИИ является одним из ключевых вызовов времени. На основании мирового и российского опыта разработки и применения искусственного интеллекта сделана первая попытка формулирования принципов доверия к медицинским приложениям в этой области.

ИСТОЧНИКИ

1. *Тополь Э.* Искусственный интеллект в медицине: Как умные технологии меняют подход к лечению. – Пер. с англ. – М.: Интеллектуальная литература, 2021. – 434 с. ISBN 978-5-9614-2920-6.
3. *Кобринский Б.А.* Системы искусственного интеллекта в медицинской практике: состояние и перспективы. – Вестник Росздравнадзора. – № 3, 2020. – С. 37–43.
4. *Финн В.К.* Искусственный интеллект: Методология, применения, философия. – М.: ЛЕНАНД, 2021. – 468 с. ISBN 978-5-9710-6266-0.
13. *Кобринский Б.А.* Искусственный интеллект в медицине: состояние и горячие точки // XIX Национальная конференция по искусственному интеллекту с международным участием КИИ-2021 (11–16 окт. 2021 г.): Тр. конф. / под ред. В.В. Борисова, Б.А. Кобринского. – Ростов-на-Дону; Таганрог: Изд-во Южного федерального ун-та, 2021. – С.13–29.

REFERENCES

1. *Topol' E.* Iskusstvennyy intellekt v meditsine: kak umnyye tekhnologii menuyut podkhod k cheloveku. – Per.s angl. – М.: Intellektual'naya literatura, 2021. – 434 s. ISBN 978-5-9614-2920-6. (in Russian).
2. *Basu K., Sinha R., Ong A., Basu T.* Artificial Intelligence: How is It Changing Medical Sciences and Its Future? // Indian J Dermatol. 2020 Sep-Oct; 65(5): 365–370. doi:10.4103/ijd.IJD_421_20.
3. *Kobrin'skii B. A.* Artificial intelligence systems in medical practice: condition and prospects. – Vestnik Roszdravnadzora – № 3, 2020. – С. 37–43. (In Russian).
4. *Finn V.K.* Iskusstvennyy intellekt: Metodologiya, primeneniye, filosofiya. – М.: LENAND, 2021. – 468 s. ISBN 978-5-9710-6266-0 (In Russian).

5. *Esteva A., Chou K., Yeung S.* et al. Deep learning-enabled medical computer vision npj Digital Medicine (2021) 4: Article number 5. doi:10.1038/s41746-020-00376-2.
6. Accenture Digital Health Technology Vision 2020. Available at: <https://www.accenture.com/us-en/insights/health/accenture-digital-health-technology-vision-2020> (Accessed 12 April 2022).
7. Healthcare execs trust AI to automate administrative processes, advance health equity: Optum. Available at: <https://www.fiercehealthcare.com/tech/healthcare-exec-trust-ai-to-automate-administrative-processes-advance-health-equity-optum> (Accessed 12 April 2022).
8. Confidence in Artificial Intelligence Increases as Health Industry Leaders Project Faster Return on Investment, Signal Growing Need for Broad Range of AI Expertise. Available at: <https://www.optum.com/about-us/news/confidence-artificial-intelligence-increases.html> (Accessed 12 April 2022).
9. *Topol E.J.* High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019; (25): 44–56. doi:10.1038/s41591-018-0300-7.
10. *Verghese A., Shah N.H., Harrington R.A.* What this computer needs is a physician: humanism and artificial intelligence. *JAMA.* 2018;(319):19–20. doi:10.1001/jama.2017.19198.
11. *Briganti G., Le Moine O.* Artificial Intelligence in Medicine: Today and Tomorrow // *Front. Med.* 2020;(7):Article 27. doi.org/10.3389/fmed.2020.00027.
12. *Price W.N., Gerke S., Cohen I.G.* Potential liability for physicians using artificial intelligence. *JAMA.* 2019;(322):1765–6. doi:10.1001/jama.2019.15064.
13. *Kobrinский B.A.* Искусственный интеллект в медицине: состояние и перспективы // Десятилетие национальной конференции по искусственному интеллекту с привлечением КИИ-2021 (11–16 окт. 2021 г.): Тр. конф. / под ред. В.В. Борисова, В.А. Кобринского. – Ростов-на-Дону: Таганрог: Изд-во Южного федерального университета, 2021. – С.13–29. (In Russian).
14. Ethics Guidelines for Trustworthy AI. Available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.1.html> (Accessed 12 April 2022).
15. ISO/IEC TR 24028:2020. Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. Available at: <https://www.iso.org/standard/77608.html> (Accessed 12 April 2022).
16. *Mathews S.C., McShea M.J., Hanley C.L.* et al. Digital health: a path to validation. *npj Digit. Med.* 2, 38 (2019). <https://doi.org/10.1038/s41746-019-0111-3>.
17. *Larsen M.E., Huckvale K., Nicholas J.* et al. Using science to sell apps: Evaluation of mental health app store quality claims. *npj Digital Medicine.* 2, 18 (2019). <https://doi.org/10.1038/s41746-019-0093-1>.
18. *Guy Martin, Saira Ghafur, Isabella Cingolani, Joshua Symons, Dominic King, Sonal Arora* et al. The effects and preventability of 2627 patient safety incidents related to health information technology failures: a retrospective analysis of 10 years of incident reporting in England and Wales. *The Lancet Digital Health*, Volume 1, ISSUE 3, e127–e135, July 01, 2019, [https://doi.org/10.1016/S2589-7500\(19\)30057-3](https://doi.org/10.1016/S2589-7500(19)30057-3).
19. *Wynants L., Van Calster B., Collins G. S., Riley R. D., Heinze G., Schuit E.* et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal *BMJ* 2020;(369):m1328 doi:10.1136/bmj.m1328, <https://doi.org/10.1136/bmj.m1328>.
20. *Roberts M., Driggs D., Thorpe M.* et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell* 3, 199–217 (2021). <https://doi.org/10.1038/s42256-021-00307-0>.
21. *Wong A., Otlis E., Donnelly J.P., Krumm A., McCullough J., DeTroyer-Cooley O., Pestrue J., Phillips M., Konye J., Penzo C., Ghaus M., Singh K.* External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med.* 2021 Aug 1; 181(8): 1065–1070. doi: 10.1001/jamainternmed.2021.2626. Erratum in: *JAMA Intern Med.* 2021 Aug 1; 181(8): 1144. PMID: 34152373; PMCID: PMC8218233, <https://doi.org/10.1001/jamainternmed.2021.2626>.
22. AI and the potential liability issues arising from use in a clinical setting. Available at: <https://www.digitalhealth.net/2019/08/ai-and-the-potential-liability-issues-arising-from-use-in-a-clinical-setting/> (Accessed 12 April 2022).
23. Informatics association seeks tweaks to FDA's AI regulatory framework. Available at: <https://www.healthcareitnews.com/ai-powered-healthcare/informatics-association-seeks-tweaks-fda%E2%80%99s-ai-regulatory-framework> (Accessed 12 April 2022).
24. *Kundu S.* AI in medicine must be explainable. *Nature Medicine* 27, 1328 (2021). <https://doi.org/10.1038/s41591-021-01461-z>
25. *Falco G., Shneiderman B., Badger J.* et al. Governing AI safety through independent audits. *Nature Machine Intelligence* 3, 566–571 (2021). <https://doi.org/10.1038/s42256-021-00370-7>
26. *Eric Perakslis, Andrea Coravos.* Is health-care data the new blood? *The Lancet Digital Health*, Volume 1, Issue 1, May 2019, Pages e8–e9, [https://doi.org/10.1016/S2589-7500\(19\)30001-9](https://doi.org/10.1016/S2589-7500(19)30001-9)
27. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) – Discussion Paper and Request for Feedback, <https://www.regulations.gov/document/FDA-2019-N-1185-0001>.
28. *P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, David A. Broniatowski, Mark A. Przybocki.* Four Principles of Explainable Artificial Intelligence. Available at: <https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf> (Accessed 12 April 2022).
29. Artificial Intelligence and Machine Learning in Software as a Medical Device. Available at: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> (Accessed 12 April 2022).
30. The Use of Artificial Intelligence in Health Care: Trustworthiness (ANSI/CTA-2090). Available at: <https://shop.cta.tech/collections/standards/products/the-use-of-artificial-intelligence-in-healthcare-trustworthiness-cta-2090> (Accessed 12 April 2022).
31. Horizon Scanning Assessment Report – Artificial Intelligence. Available at: http://www.icma.info/drupal/sites/default/files/2021-08/horizon_scanning_report_artificial_intelligence.pdf (Accessed 12 April 2022).
32. Ethics and governance of artificial intelligence for health. Available at: <https://www.who.int/publications/i/item/9789240029200> (Accessed 12 April 2022).
33. Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. Available at: <https://www.gao.gov/products/gao-21-519sp> (Accessed 12 April 2022).
34. Good Machine Learning Practice for Medical Device Development: Guiding Principles. Available at: <https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles> (Accessed 12 April 2022).