ОБЗОР ЛИТЕРАТУРЫ
LITERATURE REVIEW

УДК: 004.891
DOI: 10.37489/2782-3784-myrwd-107
EDN: WPRZIX

Качество данных реальной клинической практики как условие переносимости медицинских моделей искусственного интеллекта

Абдулкеримова С. Ф., Данилова Ю. А.
ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)», Москва, Российская Федерация

Аннотация

Актуальность. Цифровые решения, показавшие высокую точность на подготовленных исследовательских данных, всё чаще предлагаются для применения в медицинских организациях, где данные формируются под влиянием клинической практики учреждения.

Цель. Проанализировать факторы, связанные с данными реальной клинической практики, которые могут приводить к снижению качества медицинских моделей искусственного интеллекта (ИИ) при их переносе в новые условия, и обосновать подходы к их предварительной проверке перед внедрением в медицинской организации.

Материалы и методы. Выполнен обзор предметного поля с систематизацией публикаций о применимости медицинских ИИ-моделей за пределами исходной выборки. Поиск проводился в Scopus, Web of Science Core Collection, PubMed, Google Scholar и eLIBRARY. RU в период с 2016–2026 гг.; при необходимости включались более ранние источники.

Результаты. Выделены основные форматы медицинских данных, причины снижения качества моделей и подходы к оценке их применимости в новой клинической среде, и особое внимание уделено сдвигу данных как причине снижения точности, нарушения калибровки и роста числа клинически значимых ошибок. Рассмотрены подходы к внешней и локальной валидации, оценке устойчивости модели в подгруппах и мониторингу после внедрения. Также предложены практические требования практические требования к описанию их применимости в научной публикации или отчёте о валидации.

Заключение. Способность медицинской ИИ-системы сохранять эффективность при использовании вне исходной среды должна оцениваться как отдельный критерий её надёжности: в отчётах о валидации требуется обеспечивать детализированное и воспроизводимое описание источников данных, методов определения исходов, условий проведения испытаний и полученных результатов в конкретной медицинской организации, а также обоснованное представление стратегии последующего мониторинга и поддержания качества функционирования системы.

Ключевые слова: искусственный интеллект; электронные медицинские карты; качество данных; прогностические модели; внешняя валидация; локальная валидация; калибровка; клинические исходы; поддержка врачебных решений; мониторинг

Для цитирования: Абдулкеримова С. Ф., Данилова Ю. А. Качество данных реальной клинической практики как условие переносимости медицинских моделей искусственного интеллекта. *Реальная клиническая практика: данные и доказательства*. 2026;6(3):5–14. https://doi.org/10.37489/2782-3784-myrwd-107. EDN: WPRZIX.

Поступила: 10.06.2026. **В доработанном виде:** 15.07.2026. **Принята к печати:** 23.07.2026. **Опубликована:** 30.07.2026.

Quality of real-world data as a condition for the transferability of medical artificial intelligence models

Selimat F. Abdulkerimova, Yulia A. Danilova
Moscow Institute of Physics and Technology, Moscow, Russian Federation

Abstract

Relevance. The relevance of the topic is determined by the fact that digital solutions that have demonstrated high accuracy on curated research datasets are increasingly being proposed for use in healthcare organizations, where data are shaped by the clinical practices of a particular institution.

Objective. To analyze factors related to real-world data (RWD) that may lead to a decline in the performance of medical artificial intelligence (AI) models when they are transferred to new settings, and to substantiate approaches to their preliminary assessment before implementation in a healthcare organization.

Materials and methods. A scoping review was conducted to systematize publications on the applicability of medical AI models beyond the original dataset. The search was performed in Scopus, Web of Science Core Collection, PubMed, Google Scholar and eLIBRARY. RU for the period from 2016 to 2026; earlier sources were included when necessary.

Results. The review identified the main types of medical data, causes of model performance degradation and approaches to assessing model applicability in a new clinical setting. Particular attention was paid to data shift as a cause of reduced accuracy, impaired calibration and an increased number of clinically significant errors. Approaches to external and local validation, assessment of model robustness in subgroups and post-implementation monitoring were considered. Practical requirements were also proposed for describing model applicability in a scientific publication or validation report.

Conclusion. The ability of a medical AI system to maintain its effectiveness when used outside the original environment should be assessed as a separate criterion of its reliability. Validation reports should provide a detailed and reproducible description of data sources, outcome definitions, testing conditions and results obtained in a specific healthcare organization, as well as a justified strategy for subsequent monitoring and maintenance of system performance.

Keywords: artificial intelligence; electronic health records; data quality; prediction models; external validation; local validation; calibration; clinical outcomes; clinical decision support; monitoring

For citation: Abdulkerimova SF, Danilova YuA. Quality of real-world data as a condition for the transferability of medical artificial intelligence models. *Real-World Data & Evidence.* 2026;6(3):5–14. https://doi.org/10.37489/2782-3784-myrwd-107. EDN: WPRZIX.

Received: 10.06.2026. **Revision received:** 15.07.2026. **Accepted:** 23.07.2026. **Published:** 30.07.2026.

Введение / Introduction

Данные реальной клинической практики (real-world data; RWD) всё активнее используют при оценке технологий здравоохранения (ОТЗ) и анализе исходов лечения (outcome research), поскольку они позволяют рассматривать медицинскую помощь в условиях, близких к повседневной работе клиники. Их значение определяется тем, что они отражают весь путь пациента от обращения за помощью до фиксации результата лечения. В журнале «Реальная клиническая практика: данные и доказательства» уже рассматривались электронные медицинские карты как источник таких данных, подходы к оценке их качества и особенности работы с медицинской документацией в современном здравоохранении [1–3].

На этом фоне медицинские модели искусственного интеллекта (ИИ) следует рассматривать как цифровые инструменты, качество которых зависит от того, как клиническая информация была получена, записана и подготовлена к анализу. Ранее опубликованные работы о применении машинного обучения в диагностике орфанных заболеваний и о моделях поддержки врачебных решений позволяют обсуждать ИИ в логике доказательств, получаемых на данных реальной клинической практики [4, 5]. При такой постановке вопрос смещается от описания самой технологии к условиям, в которых модель получает входные данные, формирует результат и встраивается в медицинское решение.

Медицинская ИИ-модель обрабатывает цифровое представление клинического случая как такового. В зависимости от задачи этим представлением становятся записи электронной медицинской карты, диагностические изображения, текстовые заключения или данные, возникающие при работе системы поддержки врачебных решений (СППВР). Каждый формат формируется по собственным правилам, поэтому модель, успешно проверенная на одном наборе данных, при переносе в другую медицинскую организацию начинает работать в изменившейся информационной среде.

В настоящей статье сдвиг данных рассматривается как изменение условий, при которых модель получает входную информацию и сопоставляет её с клиническим исходом [6]. Этот термин можно объяснить через перенос модели из одной клиники в другую: меняется состав пациентов, порядок обследования и практика ведения медицинской документации, вследствие чего прежняя связь между цифровыми признаками и клиническим событием ослабевает. В такой ситуации прогноз может хуже соответствовать реальному риску, хотя сама модель технически продолжает работать: метрика, полученная на подготовленном исследовательском наборе данных, характеризует модель в условиях разработки, тогда как клиническое использование требует проверки в той среде, где модель должна помогать врачу. Технические обзоры по адаптации и обобщению моделей описывают причины различий между наборами данных, а для медицинской практики эти различия приобретают значение через локальную применимость, безопасность и устойчивость результата [7, 8].

В литературе уже описаны ситуации, где перенос медицинской ИИ-модели в новую клиническую среду может сопровождаться снижением качества,

нарушением калибровки и изменением практической полезности [9–17]. Эти данные обосновывают необходимость внешней и локальной валидации, а также клинической интерпретации ошибок и последующего мониторинга.

Цель / Objective

Цель обзора — определить, как свойства данных реальной клинической практики влияют на переносимость медицинских моделей искусственного интеллекта, и сформулировать требования к оценке таких моделей перед использованием в медицинской организации.

Материалы и методы / Materials and methods

Работа выполнена как обзорное исследование с систематизацией публикаций о применимости медицинских ИИ-моделей за пределами исходной выборки. Такой формат был выбран в связи с методологической неоднородностью темы, где переносимость медицинских моделей искусственного интеллекта изучается в разных областях клинического ИИ, и каждая из них по-своему связывает качество данных с применимостью модели в новой медицинской организации.

Обзорный вопрос был сформулирован следующим образом: какие свойства данных реальной клинической практики влияют на переносимость медицинских ИИ-моделей и какие подходы используются для оценки их применимости в новой медицинской организации?

Поиск проводился в PubMed, Scopus, Web of Science, Google Scholar, eLIBRARY и CyberLeninka. Рассматривались публикации 2016–2026 гг.; более ранние источники включались при наличии методологического значения для темы. Поисковая стратегия объединяла термины, относящиеся к данным реальной клинической практики, медицинскому искусственному интеллекту, электронным медицинским картам, сдвигу данных, внешней валидации,

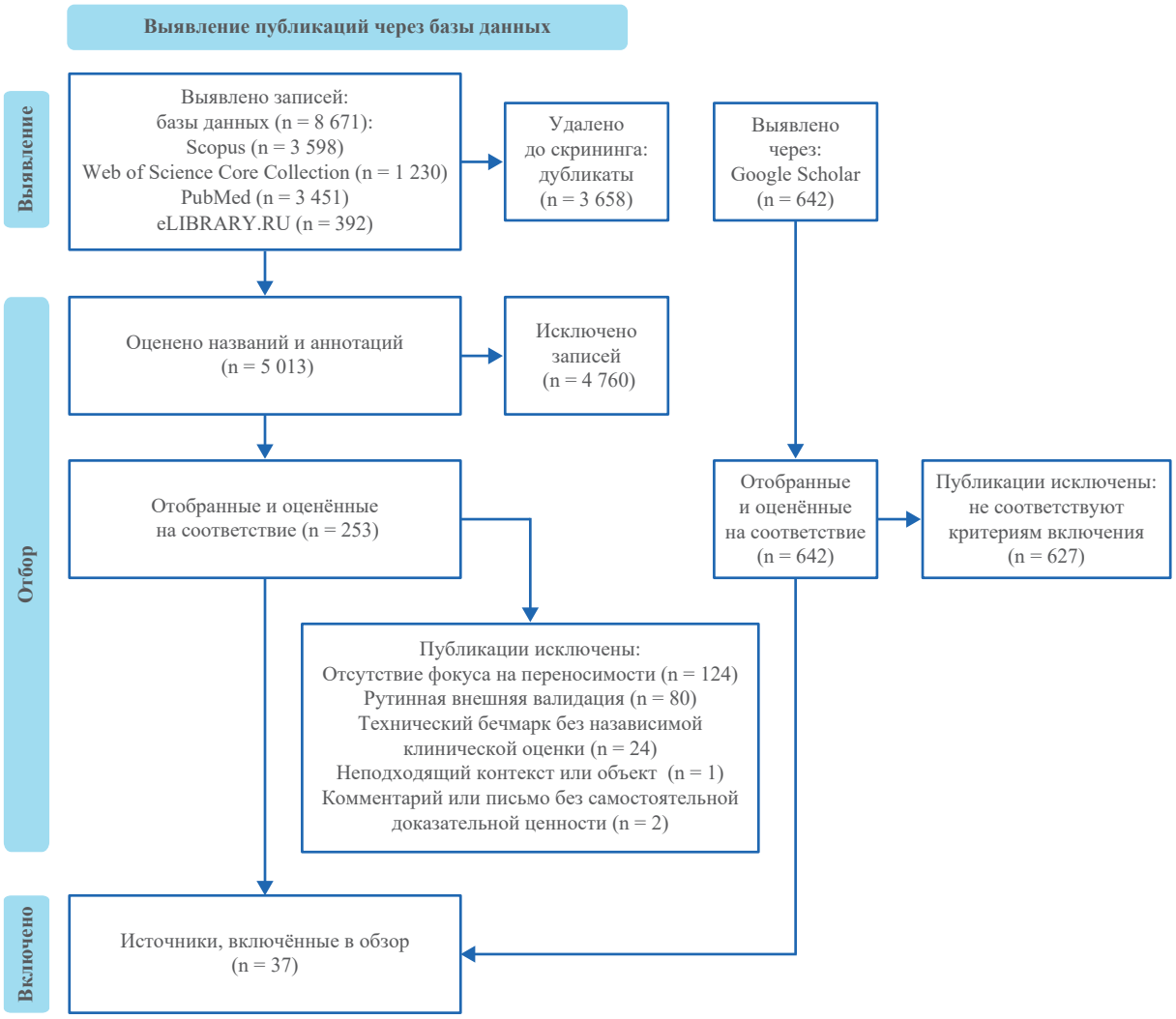

Рис. 1. *Схема отбора источников для обзора по PRISMA-ScR*
Fig. 1. *PRISMA-ScR source selection flow chart for the review*

локальной валидации, калибровке и мониторингу после внедрения.

В обзор включались статьи и методические документы, где переносимость модели рассматривалась через качество данных, внешнюю проверку, локальную валидацию, анализ ошибок или мониторинг. Публикации, посвящённые отдельным клиническим применениям ИИ без анализа переносимости, а также работы с оценкой качества только на внутреннем случайном разделении данных, в основной синтез не включались. Итоговые источники группировались по этапам перехода модели от разработки к применению: формирование данных, нарушение сопоставимости, интерпретация ошибок, мониторинг и описание результатов в публикации или отчёте.

При подготовке раздела учитывались принципы отчётности для обзоров предметного поля (scoping review), отражённые в PRISMA-ScR, и методологические рекомендации JBI [18, 19]. Протокол обзора и журнал поиска были опубликованы в Open Science Framework: doi:10.17605/OSF. IO/MDKUW [38] и включают обзорный вопрос, критерии включения и исключения, поисковые стратегии, порядок дедупликации, этапы отбора источников, план извлечения данных и способ тематической систематизации результатов. Отбор публикаций по названию и аннотации, а затем окончательная оценка соответствия выполнялись двумя исследователями независимо. Разногласия разрешались обсуждением; при сохранении расхождений решение принималось с учётом обзорного вопроса и критериев включения.

Процесс отбора источников представлен на рис. 1. В схеме отдельно показаны основная поисковая ветвь по библиографическим базам данных и дополнительный структурированный поиск в Google Scholar. Сводные причины исключения публикаций на этапе окончательной оценки соответствия приведены в протоколе обзора.

Результаты / Results

Данные реальной клинической практики как среда применения медицинской ИИ-модели. Для оценки переносимости важно установить, что именно отражает признак, попавший в аналитический набор через конкретное клиническое или организационное действие, поскольку один и тот же показатель в разных случаях имеет разный вес/вклад в прогноз. Например, если компьютерную томографию (КТ) выполнять всем подряд, признак «наличие КТ» перестаёт отражать клиническое подозрение и теряет прогностическую ценность, поскольку перестаёт различать пациентов по риску. Таким образом, порядок, который делает данные клинически содержательными, связывает их с локальной практикой учреждения, что требует оценки пригодности данных перед использованием в исследовании или валидации [1–3]. Для моделей, работающих с электронными медицинскими картами (ЭМК), особое значение имеет способ определения исхода. Клиническое событие может быть задано кодом диагноза, лабораторным результатом, назначением лечения или формализованным правилом, и каждый вариант по-своему отделяет случай от его отсутствия. При прогнозировании сепсиса это означает, что модель будет обучаться на разных клинических ситуациях в зависимости от того, какой способ фиксации события принят в исходном наборе данных; при переносе в другую организацию это различие способно изменить результат проверки.

Полнота данных также зависит от практики обследования: лабораторный показатель появляется в карте после решения врача назначить исследование, а отсутствие значения может быть связано с тяжестью состояния или доступностью обследования. В такой ситуации модель способна учитывать сам факт назначения анализа как признак риска, хотя по клиническому смыслу этот признак описывает действие врача. Для прогноза, построенного на динамике стационарного наблюдения, подобная зависимость критична, поскольку порядок накопления записей начинает влиять на прогноз.

В медицинской визуализации переносимость модели зависит от того, как в учреждении формируется сама база изображений. На исследование направляют пациентов с конкретным состоянием, поэтому набор снимков отражает практику отбора и этап диагностики. При обучении на таких данных цифровой инструмент может усвоить связь между изображением и локальной практикой обследования, а при переносе в другую клинику эта связь ослабевает. Дополнительный вклад вносит техника получения изображения: различия протоколов и оборудования способны создавать признаки, полезные в исходной базе, но нестабильные при внешней проверке [9, 11]. Таким образом, сдвиг данных в медицинской ИИ-модели возникает тогда, когда меняется путь от клинического состояния пациента к его цифровому представлению [6]. В новой среде модель продолжает использовать связи, найденные в данных разработки, но эти связи могут хуже соответствовать пациентам, исходам и практике конкретной организации, поэтому уже на этапе первичной проверки следует оценивать, какие элементы локальной практики вошли в данные разработки и сохраняются ли они в условиях будущего применения.

Примеры снижения качества моделей при переносе. Клинические примеры непереносимости медицинских ИИ-моделей наиболее

убедительно описаны в работах, где алгоритмы проверяли на данных из разных источников или в группах пациентов с различной представленностью в исходном наборе. Наиболее наглядно это видно в медицинской визуализации: в исследовании по выявлению пневмонии на рентгенограммах грудной клетки модель проверяли между несколькими больничными системами, и результат оказался различным в зависимости от того, из какого учреждения поступали изображения [9]. В исследованиях по выявлению COVID-19 на рентгенограммах грудной клетки была описана более жёсткая форма той же проблемы: модель использовала косвенные признаки условий получения изображения, и клинический сигнал смешивался с техническим и организационным контекстом. Такой механизм называют обучением на косвенных признаках (shortcut learning). Он опасен тем, что модель демонстрирует высокую точность в исходной базе, но при внешней проверке теряет устойчивость, поскольку технические и организационные особенности набора данных перестают совпадать с новой клинической средой [10]. Сходную логику подтверждают работы по переносу моделей между данными магнитно-резонансной томографии (МРТ) из разных источников, где различия в происхождении изображений требовали адаптации модели перед применением [11].

Проблема переносимости проявляется также через состав пациентов, на которых модель была разработана и проверена. В дерматологическом исследовании на специально собранном разнообразном наборе изображений современные модели показали ограничения, особенно для тёмных оттенков кожи и редких заболеваний [12]. В исследованиях алгоритмов анализа рентгенограмм грудной клетки сходная проблема описана как риск недодиагностики в недостаточно представленных группах пациентов (algorithmic underdiagnosis): модель чаще относила больных пациентов к категории «патология не выявлена», когда это были женщины, чернокожие и латиноамериканские пациенты, а также пациенты младшего возраста и пациенты со страховкой Medicaid [13]. Эти исследования доказывают необходимость локальной валидации, поскольку средняя метрика качества модели по всей выборке может скрывать клинически значимые ошибки в отдельных группах, и проверка применимости модели должна учитывать состав пациентов конкретной медицинской организации.

Отдельный риск возникает, когда целевой исход задаётся удобным для анализа, но клинически неполным косвенным показателем. В исследовании алгоритма управления здоровьем населения риск пациента оценивали через расходы на лечение [14]: предполагалось, что чем выше затраты, тем больше пациент нуждается в наблюдении. Однако расходы отражают ещё и доступность медицинской помощи, частоту обращений и сложившуюся практику ведения пациента. Поэтому у человека с тяжёлым заболеванием, который наблюдается реже или получает необходимую помощь позже, модель могла занижать потребность в ней.

Для моделей, работающих с электронными медицинскими картами, показательным остается пример широко внедренной модели прогнозирования сепсиса Epic Sepsis Model. При проверке в независимой больничной выборке система недостаточно хорошо отделяла пациентов с последующим сепсисом от пациентов без этого исхода, а рассчитанный риск не всегда соответствовал фактической частоте события [15]. В практическом отношении это означало, что предупреждения могли появляться у большого числа пациентов, но далеко не каждое из них указывало на клинически подтверждаемый риск, поэтому нагрузка на персонал возрастала без сопоставимого повышения надёжности раннего выявления сепсиса. Исследования внедрения Sepsis Watch дополняют эту картину, показывая, что после внедрения польза/эффективность модели зависит от того, своевременно ли предупреждение поступает ответственному специалисту и понимает ли врач, что именно нужно проверить или изменить в ведении пациента [16]. Концепция медицинского алгоритмического аудита (medical algorithmic audit) развивает эту логику, предлагая оценивать модель через ошибки, возможный вред и условия фактического применения [17].

Основные примеры, на которых основано обсуждение и которые иллюстрируют разные механизмы снижения качества при переносе модели, обобщены в табл. 1.

Оценка применимости моделей и минимизация риска ошибок. За пределами исходного набора данных модель может сохранять приемлемую метрику, однако решение о её применении в учреждении требует локальной валидации, на этапе которой устанавливают, работает ли прогноз на данных будущего применения и требуется ли адаптация модели перед внедрением [20–22].

Клиническая интерпретация прогностической модели требует оценки калибровки, то есть проверки того, насколько числовой прогноз модели соответствует фактически наблюдаемой частоте исхода. Хорошо калиброванная модель помимо того, что распределяет пациентов по уровню риска, выдаёт вероятность, которую можно использовать при выборе клинического действия. Площадь под ROC-кривой показывает, как модель ранжирует пациентов, и помогает сравнивать пациентов с разным уровнем риска, однако такой показатель

Таблица 1. Показательные исследования, демонстрирующие причины снижения качества медицинских ИИ-моделей при переносе
Table 1. Representative studies demonstrating the reasons for the decline in the quality of medical AI models during transfer

Авторы	Статья	Клиническая область и данные	Модель	Что показано в исследовании	Методологический вывод для оценки переносимости
Zech et al., 2018 [9]	Выявление пневмонии на рентгенограммах грудной клетки при переносе между больничными системами [9]	Рентгенография грудной клетки; данные нескольких больничных систем	Свёрточная нейронная сеть для выявления пневмонии на рентгенограммах грудной клетки	Качество модели выявления пневмонии различалось при переносе между учреждениями	Внешняя проверка должна выполняться на данных, отличающихся от исходного источника
DeGrave et al., 2021 [10]	Выявление COVID-19 на рентгенограммах грудной клетки и обучение на косвенных признаках (shortcut learning) [10]	Рентгенография грудной клетки при COVID-19; данные, полученные в разных условиях	Глубокие модели классификации COVID-19 по рентгенограммам грудной клетки	Модель использовала косвенные признаки, связанные с условиями получения изображения	Требуется анализ того, какие признаки фактически использует модель
Ghafoorian et al., 2017 [11]	Перенос модели сегментации поражений между МРТ-данными из разных источников [11]	МРТ для сегментации поражений головного мозга; данные из разных центров и при разных протоколах	CNN-модель сегментации с переносом обучения и дообучением на целевом домене	Различия между источниками МРТ-данных влияли на переносимость модели	Для изображений требуется проверка устойчивости к различиям источников и протоколов
Daneshjou et al., 2022 [12]	Оценка дерматологических ИИ-моделей на разнообразном клиническом наборе изображений [12]	Дерматологические изображения; пациенты с разными оттенками кожи и заболеваниями	Дерматологические ИИ-модели для различения доброкачественных и злокачественных поражений кожи	Качество различалось у пациентов, недостаточно представленных в наборе данных	Усредненная метрика оценки качества должна дополняться отдельной оценкой в клинически значимых группах
Seyyed-Kalantari et al., 2021 [13]	Недодиагностика в алгоритмах анализа рентгенограмм грудной клетки (algorithmic underdiagnosis) [13]	Рентгенография грудной клетки; группы пациентов с разной представленностью в данных	Модели компьютерного зрения для классификации патологий на рентгенограммах грудной клетки	Ошибки распределялись неравномерно между группами пациентов	Локальная проверка должна учитывать состав пациентов и риск неравномерных ошибок
Obermeyer et al., 2019 [14]	Использование расходов на медицинскую помощь как косвенного показателя медицинской потребности [14]	Административные данные в управлении здоровьем населения	Коммерческий алгоритм стратификации риска для программ управления здоровьем населения	Косвенный показатель исказил клинический смысл риска	Перед обучением и валидацией требуется проверять, что целевая переменная соответствует медицинской задаче
Wong et al., 2021 [15]	Внешняя проверка внедренной модели прогнозирования сепсиса по данным ЭМК [15]	Электронные медицинские карты; прогнозирование сепсиса	Epic Sepsis Model — проприетарная модель раннего предупреждения сепсиса	Внешняя проверка выявила ограничения дискриминации и калибровки внедренной модели	Широкое внедрение модели требует независимой проверки качества в конкретной медицинской организации
Sendak et al., 2020 [16]	Внедрение системы раннего предупреждения сепсиса в клиническую работу [16]	Система раннего предупреждения сепсиса	Sepsis Watch — платформа на основе глубокого обучения, интегрированная в клинический рабочий процесс	Эффект модели зависел от организации её использования	Оценка модели должна учитывать действия, которые следуют после предупреждения
Liu et al., 2022 [17]	Медицинский алгоритмический аудит (medical algorithmic audit) [17]	Методология оценки медицинских алгоритмов	Медицинский алгоритмический аудит как рамка оценки ошибок, вреда и условий применения	Модель предлагается оценивать через ошибки, возможный вред и условия применения	Для клинического использования необходим алгоритмический аудит

не отвечает на вопрос, насколько численное значение прогноза пригодно для принятия решения. При переносе модели в другую медицинскую организацию это соответствие может нарушаться, поскольку меняется состав пациентов, частота целевого события и практика лечения [23].

Когда рассчитанный риск должен приводить к определённым действиям врача, заранее задают порог, после которого результат модели влияет на клиническое решение. Например, в диагностической задаче может определяться необходимость дополнительного обследования, а в системе раннего предупреждения — момент информирования медицинского персонала. Анализ кривой принятия решений (decision curve analysis; DCA) позволяет сопоставить положительный эффект от предупреждений с последствиями ложных и пропущенных срабатываний [24], а при редких исходах особое значение имеет частота подтверждённых случаев среди всех предупреждений, поскольку именно она показывает, насколько работа системы будет приемлема для медицинской организации [25]. После внедрения эта оценка должна продолжаться: клиническая практика, правила документирования и цифровая среда учреждения меняются, врач начинает действовать с учётом предупреждений модели, и поэтому требуется мониторинг, который позволяет выявлять изменение качества прогноза и своевременно пересматривать условия применения системы [26, 27].

Наконец, описывать валидацию следует так, чтобы читатель мог восстановить условия, в которых была получена метрика. В статье или отчёте необходимо показать, где и на каких данных проверялась модель, как определялся клинический исход и в каком сценарии предполагается использовать результат. Современные руководства по отчётности для прогностических, диагностических и интервенционных исследований с применением ИИ требуют подробно описывать данные, модель, условия проверки и клинический сценарий применения [28–35]. В исследованиях на данных реальной клинической практики необходимо указывать происхождение данных и клинический смысл прогноза, поскольку именно эти элементы определяют, что фактически предсказывает модель и в каких условиях её результат может считаться применимым.

Значение более понятного и прозрачного отчёта подтверждается опубликованным анализом систем с применением ИИ, допущенных FDA к медицинскому применению, который показывает, что в материалах о таких продуктах часто недостаточно подробно описаны данные, на которых их проверяли, и условия выполненной валидации [36]. Поэтому числовые показатели качества следует связывать с клиническим решением: ROC-анализ, чувствительность и специфичность помогают интерпретировать результат модели только тогда, когда понятно, для какой диагностической или прогностической задачи они рассчитаны [37].

Пробелы в литературе и направления будущих исследований. В литературе сохраняется разрыв между требованиями к отчётности медицинских ИИ-моделей и воспроизводимыми процедурами их локальной проверки в конкретной медицинской организации. Руководства DECIDE-AI, TRIPOD+AI и другие стандарты повышают прозрачность описания валидации и ранней клинической оценки ИИ-систем, однако не задают единый порядок решения о применимости модели на локальных данных [28–35]. Методологическая неполнота особенно заметна для калибровки — она часто не представлена в публикациях и это ограничивает интерпретацию вероятностных прогнозов, а также затрудняет выбор клинических порогов при переносе модели в новую среду, так как проверка (будь то локальная или внешняя) ограничена AUC, чувствительностью и специфичностью, не позволяет полноценно оценить пригодность вероятностной модели для практического решения. Так как актуальность данной проблемы подтверждается позицией регуляторов, дальнейшие исследования должны быть направлены на разработку стандартов локальной валидации и мониторинга. В перспективе требуется интеграция методов непрерывного мониторинга с инфраструктурой электронных медицинских записей, разработки стандартов отчётности для наблюдений после внедрения ИИ-систем и создания общедоступных реестров эффективности моделей в реальной практике. Дополнительно важным направлением является изучение взаимодействия человека и алгоритма, включая влияние доверия к системе на клинические решения, а также разработка адаптивных моделей, способных обновляться без потери воспроизводимости и прозрачности.

Обсуждение и заключение / Discussion and conclusion

Проведённый обзор литературы показывает, что переносимость медицинской ИИ-модели следует рассматривать как самостоятельное условие её доказательности при использовании данных реальной клинической практики, поскольку результат, полученный на подготовленном исследовательском наборе, описывает среду разработки, а клиническое применение требует проверки там, где прогноз должен влиять на медицинское решение. Если меняется путь от состояния пациента к его цифровому представлению, цифровой инструмент

может сохранять техническую работоспособность и одновременно хуже отражать реальный риск. Поэтому вывод о применимости должен опираться на локальную проверку и последующее наблюдение за качеством после внедрения.

Ограничения исследования / Study limitations

Настоящий обзор имеет ограничения, требующие упоминания. В обзор включались преимущественно англоязычные и русскоязычные источники, поэтому публикации на других языках могли остаться вне анализа. Дополнительным ограничением может являться и то, что формальная оценка риска систематической ошибки не проводилась, поскольку задача обзора предметного поля состояла в систематизации механизмов нарушения переносимости, а не в количественном сравнении эффективности моделей. Кроме того, выбор «показательных» публикаций носит неизбежно субъективный характер: при большом числе доступных исследований в таблицу включались работы, которые, по мнению авторов, наиболее наглядно иллюстрируют ключевые механизмы снижения переносимости.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов в связи с данной публикацией.

Участие авторов

Все авторы внесли значимый вклад в проведение исследования и подготовку статьи, прочли и одобрили финальную версию статьи перед публикацией, выразили согласие нести ответственность за все аспекты работы, подразумевающую надлежащее изучение и решение вопросов, связанных с точностью или добросовестностью любой части работы. Абдулкеримова С. Ф. — концепция работы, поиск литературы, анализ источников, подготовка текста рукописи, Данилова Ю. А. — научное редактирование, критический пересмотр содержания рукописи.

Финансирование

Исследование выполнено без целевого внешнего финансирования.

Доступность данных

Первичные данные пациентов в работе не использовались, обзор основан на опубликованных источниках.

ADDITIONAL INFORMATION

Conflict of interests

The authors declare that there is no conflict of interest in connection with this publication.

Participation of authors

All authors made a significant contribution to the research and preparation of the manuscript, read and approved the final version of the article before publication, and agreed to be responsible for all aspects of the work, including proper study and resolution of issues related to the accuracy or integrity of any part of the work. Abdulkerimova S. F. — concept of the work, literature search, source analysis, preparation of the manuscript text, Danilova Yu. A. — scientific editing, critical revision of the manuscript content.

Financing

The study was carried out without targeted external funding.

Data availability

Primary patient data were not used in this study; the review is based on published sources.

СВЕДЕНИЯ ОБ АВТОРАХ

Абдулкеримова Селимат Феликсовна — магистр, ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)», Москва, Российская Федерация

Автор, ответственный за переписку

e-mail: atgcubuilder@mail.ru

ORCID ID: 0009-0009-7568-8069

Данилова Юлия Андреевна — магистр, ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)», Москва, Российская Федерация

ORCID ID: 0009-0000-2640-7472

ABOUT THE AUTHORS

Selimat F. Abdulkerimova — Master's degree, Moscow Institute of Physics and Technology (National Research University), Moscow, Russian Federation

Corresponding author

e-mail: atgcubuilder@mail.ru

ORCID ID: 0009-0009-7568-8069

Yulia A. Danilova — Master's degree, Moscow Institute of Physics and Technology (National Research University), Moscow, Russian Federation

ORCID ID: 0009-0000-2640-7472

Литература / References

1. Гусев А.В., Зингерман Б.В., Тюфилин Д.С., Зинченко В.В. Электронные медицинские карты как источник данных реальной клинической практики. *Реальная клиническая практика: данные и доказательства*. 2022;2(2):8-20. DOI: 10.37489/2782-3784-myrwd-13. EDN: HMDMZY [Gusev A.V., Zingerman B.V., Tyufilin D.S., Zinchenko V.V. Electronic medical records as a source of real-world clinical data. *Real-World Data & Evidence*. 2022;2(2):8-20. (In Russ.)].
2. Боровская В.Г., Гомон Ю.М. Оценка качества данных реальной клинической практики. *Реальная клиническая практика: данные и доказательства*. 2022;2(4):10-16. DOI: 10.37489/2782-3784-myrwd-22. EDN: CKCSIS [Borovskaya V.G., Gomon Y.M. Quality assessment of real-world data. *Real-World Data & Evidence*. 2022;2(4):10-16. (In Russ.)].
3. Свечкарева И.Р., Курылев А.А., Шилова Д.Е. Особенности оценки данных электронных медицинских карт в современном здравоохранении. *Реальная клиническая практика: данные и доказательства*. 2024;4(4):28-34. DOI: 10.37489/2782-3784-myrwd-060. EDN: YJFMTQ [Svechkareva I.R., Kurylev A.A., Shilova D.E. Particular qualities of evaluation of electronic card data in modern healthcare. *Real-World Data & Evidence*. 2024;4(4):28-34. (In Russ.)].
4. Дмитриева Н.Ю. Возможности машинного обучения для диагностики орфанных заболеваний. *Реальная клиническая практика: данные и доказательства*. 2023;3(3):36-39. DOI: 10.37489/2782-3784-myrwd-40. EDN: XJXWRQ [Dmitrieva N.Y. Machine learning capabilities for the diagnosis of orphan diseases. *Real-World Data & Evidence*. 2023;3(3):36-39. (In Russ.)].
5. Грибова В.В., Окунь Д.Б., Шалфеева Е.А. Модель поддержки врачебных решений в диагностике и реабилитации пациентов с ограничениями жизнедеятельности. *Реальная клиническая практика: данные и доказательства*. 2025;5(4):81-96. DOI: 10.37489/2782-3784-myrwd-091. EDN: ZSOAYM [Gribova V.V., Okun D.B., Shalfeeva E.A. Medical decision support model for diagnosing and rehabilitating patients with disabilities. *Real-World Data & Evidence*. 2025;5(4):81-96. (In Russ.)].
6. Finlayson SG, Subbaswamy A, Singh K, et al. The Clinician and Dataset Shift in Artificial Intelligence. *The New England Journal of Medicine*. 2021 Jul;385(3):283-286. DOI: 10.1056/nejmc2104626.
7. Guan H, Guan H, Liu M. Domain Adaptation for Medical Image Analysis: A Survey. *IEEE Transactions on Bio-medical Engineering*. 2022 Mar;69(3):1173-1185. DOI: 10.1109/tbme.2021.3117407.
8. Zhou K, Liu Z, Qiao Y, et al. Domain Generalization: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023 Apr;45(4):4396-4415. DOI: 10.1109/tpami.2022.3195549.
9. Zech JR, Badgeley MA, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *Plos Medicine*. 2018 Nov;15(11):e1002683. DOI: 10.1371/journal.pmed.1002683.
10. DeGrave AJ, Janizek JD, Lee SI. AI for radiographic COVID-19 detection selects shortcuts over signal. *medRxiv [Preprint]*. 2020 Oct 7:2020.09.13.20193565. doi: 10.1101/2020.09.13.20193565. Update in: This article has been published with doi: 10.1038/s42256-021-00338-7.
11. Ghafoorian M, Mehrtash A, Kapur T, et al. Transfer learning for domain adaptation in MRI: application in brain lesion segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2017. Cham: Springer; 2017:516–524. doi:10.1007/978-3-319-66179-7_59.
12. Daneshjou R, Vodrahalli K, Novoa RA, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci Adv*. 2022 Aug 12;8(32):eabq6147. doi: 10.1126/sciadv.abq6147.
13. Seyyed-Kalantari L, Zhang H, McDermott MBA, et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med*. 2021 Dec;27(12):2176-2182. doi: 10.1038/s41591-021-01595-0.
14. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019 Oct 25;366(6464):447-453. doi: 10.1126/science.aax2342.
15. Wong A, Otles E, Donnelly JP, et al. 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. doi: 10.1001/jamainternmed.2021.3907.
16. Sendak MP, Ratliff W, Sarro D, et al. Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study. *JMIR Med Inform*. 2020 Jul 15;8(7):e15182. doi: 10.2196/15182.
17. Liu X, Glocker B, McCradden MM, et al. The medical algorithmic audit. *Lancet Digit Health*. 2022 May;4(5):e384-e397. doi: 10.1016/S2589-7500(22)00003-6. Epub 2022 Apr 5. Erratum in: Lancet Digit Health. 2022 Jun;4(6):e405. doi: 10.1016/S2589-7500(22)00089-9.

18. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018 Oct 2;169(7):467-473. doi: 10.7326/M18-0850.
19. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth*. 2020 Oct;18(10):2119-2126. doi: 10.11124/JBIES-20-00167.
20. Youssef A, Pencina M, Thakur A, et al. External validation of AI models in health should be replaced with recurring local validation. *Nat Med*. 2023 Nov;29(11):2686-2687. doi: 10.1038/s41591-023-02540-z.
21. Collins GS, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ*. 2024 Jan 8;384:e074819. doi: 10.1136/bmj-2023-074819.
22. Riley RD, Archer L, Snell KIE, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ*. 2024 Jan 15;384:e074820. doi: 10.1136/bmj-2023-074820.
23. Van Calster B, McLernon DJ, van Smeden M, et al; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019 Dec 16;17(1):230. doi: 10.1186/s12916-019-1466-7.
24. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006 Nov-Dec;26(6):565-74. doi: 10.1177/0272989X06295361.
25. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning. 2006:233–240.
26. Koch LM, Baumgartner CF, Berens P. Distribution shift detection for the postmarket surveillance of medical AI algorithms: a retrospective simulation study. *NPJ Digit Med*. 2024 May 9;7(1):120. doi: 10.1038/s41746-024-01085-w.
27. Guo LL, Pfohl SR, Fries J, et al. Systematic Review of Approaches to Preserve Machine Learning Performance in the Presence of Temporal Dataset Shift in Clinical Medicine. *Appl Clin Inform*. 2021 Aug;12(4):808-815. doi: 10.1055/s-0041-1735184.
28. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024 Apr 16;385:e078378. doi: 10.1136/bmj-2023-078378. Erratum in: BMJ. 2024 Apr 18;385:q902. doi: 10.1136/bmj.q902.
29. Sounderajah V, Guni A, Liu X, et al; STARD-AI Steering Committee. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med*. 2025 Oct;31(10):3283-3289. doi: 10.1038/s41591-025-03953-8. Epub 2025 Sep 15. Erratum in: Nat Med. 2026 Jul 13. doi: 10.1038/s41591-026-04570-9.
30. Liu X, Cruz Rivera S, Moher D, et al; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020 Sep;26(9):1364-1374. doi: 10.1038/s41591-020-1034-x.
31. Cruz Rivera S, Liu X, Chan AW, et al; SPIRIT-AI and CONSORT-AI Working Group; SPIRIT-AI and CONSORT-AI Steering Group; SPIRIT-AI and CONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020 Sep;26(9):1351-1363. doi: 10.1038/s41591-020-1037-7.
32. Vasey B, Nagendran M, Campbell B, et al; DECIDE-AI expert group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. 2022 May 18;377:e070904. doi: 10.1136/bmj-2022-070904.
33. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol Artif Intell*. 2020 Mar 25;2(2):e200029. doi: 10.1148/ryai.2020200029.
34. Norgeot B, Quer G, Beaulieu-Jones BK, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med*. 2020 Sep;26(9):1320-1324. doi: 10.1038/s41591-020-1041-y.
35. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025 Mar 24;388:e082505. doi: 10.1136/bmj-2024-082505.
36. Wu E, Wu K, Daneshjou R, et al. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med*. 2021 Apr;27(4):582-584. doi: 10.1038/s41591-021-01312-x.
37. Григорьев С.Г., Лобзин Ю.В., Скрипченко Н.В. Роль и место логистической регрессии и ROC-анализа в решении медицинских диагностических задач. *Журнал инфектологии*. 2016;8(4):36-45. DOI: 10.22625/2072-6732-2016-8-4-36-45 [Grigoryev S.G., Lobzin Yu.V., Skripchenko N.V. The role and place of logistic regression and ROC analysis in solving medical diagnostic task. *Journal Infectology*. 2016;8(4):36-45. (In Russ.)].
38. Abdulkerimova SF. Quality of Real-world Clinical Data as a Condition for Transferability of Medical Artificial Intelligence Models: Scoping Review Protocol and Search Record. *OSF*. 2026; doi:10.17605/OSF.IO/MDKUW.