Искусственный интеллект (ИИ) в здравоохранении переживает период хайпа. У всех разработчиков масса амбиций, начиная от мелких стартапов и заканчивая крупными транснациональными вендорами и производителями медицинского оборудования, которые предлагают самые разные алгоритмы, сервисы и т.д. Однако создание решений на основе искусственного интеллекта для медицины и здравоохранения – это не только новые возможности, но и многочисленные проблемы. В этой статье рассмотрим барьеры, которые нужно преодолеть на этом пути заранее, чтобы не совершать типичные ошибки, а также ответим на главный вопрос, который волнует абсолютно всех: заменит ли ИИ врача в будущем?

All-over-IP

Большое количество команд разработчиков, давно и профессионально занимающихся системами компьютерного зрения, искусственным интеллектом и нейронными сетями, активно идут в сферу здравоохранения. Порой это специалисты с совершенно потрясающими перспективами, большим опытом и готовыми продуктами, которые успешно внедряются в промышленности, на транспорте и др. Но все они приходят в абсолютно новую для себя область, которая живет по своим собственным законам и правилам. И главное в ней то, что ценой даже за малейшую ошибку может стать жизнь человека.

Доказательная медицина
Когда речь идет о цифровизации в медицине, очень часто говорят о консервативности врачей, сопротивлении врачебного сообщества и т.д. Но так считают те, кто не понимает, каким путем здесь формируются решения, как внедряются любые технологии, будь то таблетка, новый вид скальпеля или цифровой продукт.

Если посмотреть на типовую пирамиду доказательной медицины с разными уровнями доказательности научных исследований, то, к сожалению, подавляющее количество продуктов на базе искусственного интеллекта пока расположены на ее "дне". Другими словами, это единичные описания единичных случаев. Для профессионального сообщества такая мизерная информация доказательным аргументом не является.

Это значит, что надо не только разрабатывать алгоритмы на основе ИИ, но и идти путем научной валидации, растить опыт, доходить до уровня исследований и затем – метаанализов и клинических рекомендаций.

Повышение производительности и точности скринингов
В указе президента "О развитии искусственного интеллекта в Российской Федерации" № 490 от 10.10.2019 г. есть отдельное положение о повышении качества услуг здравоохранения за счет использования технологий ИИ. Потенциально внедрение искусственного интеллекта в медицине открывает большие возможности для контроля качества и автоматизации участков производственных процессов, которые достаточно типизированы и состоят из повторения рутинных операций. Чаще всего такими процессами являются скрининги и профилактические исследования, когда однотипная стандартизированная диагностика выполняется на больших объемах пациентов. Так как это условно здоровые люди, то при массовом анализе значительная часть исследований представляет собой норму и, как следствие, идеальный вариант для автоматизации.

Глобальные исследования подтверждают данную точку зрения. Согласно опросу Европейского общества рентгенорадиологии (European Society of Radiology)1, большинство врачей отдает предпочтение использованию ИИ для скрининга онкологических и сердечно-сосудистых заболеваний.

Кто несет ответственность?
Что касается разделения ответственности врача и искусственного интеллекта, то к этому вопросу медицинские работники относятся скептически. Многие из них считают, что само по себе заключение без валидации врачом пациент вряд ли может воспринять.

Другое дело – разделение ответственности между врачом и разработчиком ИИ-решения. Ведь просто разработать алгоритм и передать его в медицину не получится. Придется нести ответственность за работу этого алгоритма и в длительной перспективе.

Кроме того, на всех этапах разработки решения для здравоохранения обязательно участие врачей, причем как специалистов в сфере клинической дисциплины, которой адресовано конкретное решение, так и специалистов по организации здравоохранения.

Качество валидации ИИ для диагностики
Основная проблема, с которой сталкиваются разработчики ИИ для медицины, – это качество и стандартизация исходных данных. На рис. 3 показаны результаты обобщающей аналитической публикации2, посвященной анализу датасетов и качеству алгоритмов на их основе, разработанных для здравоохранения за несколько предыдущих лет. Обзору уже три года, но ситуация, к сожалению, не меняется.

В большинстве случаев разработчики берут отдельно взятый датасет, чаще всего принадлежащий одной больнице, тренируют на нем алгоритм и выпускают в свет. Если цель – получить инвестиции, то такой подход позволит это сделать. Но если нужно создать продукт, который будет массово работать в системе здравоохранения в стране и за рубежом, то подход неприемлем, так как результаты работы ИИ обычно оказываются невоспроизводимыми на других датасетах, диагностических устройствах, другой популяции и т.д. То есть нет валидации алгоритмов на независимых данных при разработке, а значит предлагаемый ИИ не сможет работать "за пределами" известного ему датасета.

Далее в игру вступает доказательная медицина. Если не показать врачу метрики диагностической точности алгоритма, не назвать значения площади под кривой, чувствительности, специфичности, то команда может быть сколько угодно профессиональной, а инвестиции огромными, но никто в медицинском сообществе не будет принимать это во внимание. В медицине ИИ должен пройти научную оценку, клинические испытания, наконец – получить регистрационное свидетельство в качестве медицинского изделия. Только после этого можно рассчитывать на массовое применение алгоритма. Причем такая ситуация носит глобальный характер, в каждой развитой стране есть своя система сертификации изделий и средств для здравоохранения.

Также результаты работы алгоритма обязательно должны быть воспроизводимы и объяснимы. Недостаточно сказать врачу, что на одном снимке есть патология, а на другом ее нет. Недостаточно просто показать те области, которые вызвали у алгоритма интерес и сомнение, необходимо объяснить, почему он принял то или иное решение.

Мы сталкивались с известными ситуациями, когда алгоритм учили на компьютерных томограммах буквально находить белое на черном, и в итоге он не понимал разницу между очагом в легком и содержимым.

Когда врач видит подобную картинку, он больше не будет работать с этим алгоритмом, и здесь дело не в его косности или консерватизме, а в том, что это элементарные вещи: сначала научить алгоритм находить легкие на изображении, понимать их анатомические границы, а уже потом искать очаги. Типичные грубые ошибки происходят и тогда, когда ищут просто какие-то образования без привязки к анатомическим ориентирам, и очагом заболевания может оказаться все что угодно, вплоть до застежки на нижнем белье.

Опыт сотрудничества с разработчиками
Центр диагностики и телемедицины ДЗМ (директор Центра – д.м.н., профессор Морозов Сергей Павлович) ведет масштабную и многогранную работу с разработчиками ИИ.

В этой практике взаимодействия можно выделить несколько основных проблем.

Стабильность работы команды
Мы сталкивались с ситуациями, когда последовательно тестировали одну и ту же разработку и видели, как колеблется значение метрик диагностической точности. Это было связано с тем, что постоянно менялась команда разработчиков, между разработками не было преемственности. Поэтому в контексте клинических испытаний алгоритмов искусственного интеллекта для медицины рассматриваются не только технические испытания, но и система требований к принципам управления качеством, по которым работает команда.

Отсутствие целеполагания, необоснованная постановка задач
Многие разработчики считают, что в медицине надо просто распознать какое-то изображение, мол вон сколько в ней визуализации – рентгенограммы, томограммы и др. А зачем, например, распознавать рентгенограмму таза? В какой производственный процесс встраивается алгоритм? Кто и почему будет платить за его работу? Эти вопросы, как правило, остаются без ответа.

Разработка искусственного интеллекта для медицины начинается с обоснованного целеполагания. Она выстраивается по той же схеме, что и создание медикаментов в фармбизнесе, и включает в себя не только обязательные клинические испытания, но и постмаркетинговые наблюдения.

ИИ – медицинское изделие
Часто ведутся дискуссии о том, что искусственный интеллект не нужно сертифицировать как медицинское изделие. Но при этом, если человек планирует лечить себя или члена семьи таблетками, он хочет, чтобы таблетки были проверены в клинических испытаниях, а используемые во время операции инструменты были медицинскими изделиями, стерильными, проверенными и качественными. А как только речь заходит об искусственном интеллекте, то почему-то он не является медицинским изделием и диагностическое решение, определяющее судьбу человека, воспринимается как некое приложение без всякой ответственности. Это совершенно порочная практика.

Во всем мире наблюдается четкая практика сертификации искусственного интеллекта как медицинского изделия наравне с любыми другими видами программного обеспечения.

Онлайн-обзор "Системы видеоаналитики" - решения от производителей и мнения экспертов >>>

Московский эксперимент по применению компьютерного зрения в лучевой диагностике
В 2020 г. в соответствии с постановлением Правительства Москвы Центр диагностики и телемедицины ДЗМ начал крупнейшее в мире научное исследование применимости и качества искусственного интеллекта для здравоохранения. В нем могут принять участие разработчики ИИ-решений для медицины, у которых уже есть готовый продукт (алгоритм, сервис). Он должен уметь приоритизировать исследования, обозначать патологические участки на изображениях и помогать врачу сформировать шаблон описания, то есть контролировать качество и повышать производительность работы.

В рамках эксперимента сервисы искусственного интеллекта подключаются к Единому радиологическому информационному сервису (ЕРИС), который объединяет все государственные медицинские организации города Москвы. К ЕРИС подключены около 1,5 тыс. единиц цифровой диагностической аппаратуры, ежегодно в нем накапливаются миллионы исследований, в системе работают врачи и рентгенолаборанты.

На рис. 6 показаны первые промежуточные результаты эксперимента. Он проводится по трем видам исследований: рентгенограмма грудной клетки, компьютерная томограмма грудной клетки и маммограмма. Имеются отдельные условия для алгоритмов по диагностике COVID-19.

Это уникальная возможность для разработчиков апробировать алгоритмы в реальных клинических условиях, получить обратную связь не только от команды ученых, но и от практикующих врачей, выяснить объективные метрики качества и точности, оценить воспроизводимость работы алгоритмов, а на основе этих результатов – усовершенствовать свою разработку.

С научной точки зрения эксперимент представляет собой модель клинических испытаний, разработанную коллективом Центра под руководством профессора Сергея Морозова3.

Эффективный ИИ в медицине
За искусственным интеллектом в медицине большое будущее, хотя этот инструмент еще только формируется.