Этическая мысль
2024. Т. 24. № 1. С. 115–128
УДК 177

Ethical Thought
2024, Vol. 24, No. 1, pp. 115–128
DOI: 10.21146/2074-4870-2024-24-1-115-128

Е.В. Введенская

Цифровые агенты в медицине: новые возможности и вызовы

Введенская Елена Валерьевна – кандидат философских наук, доцент. ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России. Российская Федерация, 117321, г. Москва, ул. Островитянова, д. 1.

ORCID: 0000-0003-1100-0033
e-mail: vvedenskaya.elena@gmail.com

Предметом данной статьи являются новые возможности и вызовы, к которым приводит широкое распространение цифровых агентов в медицине. Под «цифровыми агентами» понимаются все устройства с ИИ, нейронные сети, мобильные приложения, чат-боты, используемые для диагностики и мониторинга состояния пациента. Наряду с преимуществами их внедрения существует ряд рисков, провоцируемых утратой человеческого фактора и недостатком регулирования новых технологий. В статье рассматриваются «этические ловушки», без устранения которых невозможно эффективное внедрение цифровых агентов в здравоохранение. К ним относятся: предвзятость, галлюцинации и вредная информация, конфиденциальность и согласие субъекта данных, а также прозрачность и объяснимость. Отмечено, что появляется все большее делегирование ИИ принятия решений о диагнозе и лечении, ограничивающее автономию врачей в пользу предварительно заданных алгоритмов. Такой феномен «цифрового патернализма» влечет за собой утрату доверия к квалифицированным врачам. Кроме того, в статье анализируется проблема социального расслоения, вызванная как ограничением доступа к современным технологиям для определенных слоев населения, так и созданием барьеров для людей, предпочитающих традиционные подходы к оказанию медицинской помощи. Делается вывод о необходимости тщательного обсуждения и разработки правил и нормативов для внедрения цифровых агентов в медицину, с учетом максимального уважения прав пациентов, сохранения человеческих отношений «врач–пациент» и этических ценностей.

Ключевые слова: искусственный интеллект, цифровые агенты, медицина, ChatGPT, этика, предвзятость, конфиденциальность, объяснимость, благоразумие

Введение

Современная медицина находится на передовом рубеже технологических инноваций, внедрение которых способствует переосмыслению традиционных методов диагностики, лечения и ухода за пациентами. В этом контексте ключевую роль начинают играть цифровые агенты, представляющие собой интеграцию искусственного интеллекта (ИИ), автоматизированных систем и компьютерных технологий. Цифровые агенты широко востребованы в медицинской отрасли, где помогают накапливать и сравнивать данные, заполнять и вести медицинские карты, оперативно получать клиническую информацию, давать лечебные рекомендации и решать множество других прикладных задач.

Цифровые агенты в медицине открывают новые горизонты, предоставляя уникальные возможности для повышения эффективности диагностики, персонализированного лечения и постоянного мониторинга состояния пациентов. В рамках модели партисипативной медицины пациент вовлекается в процесс лечения как активный участник, и здесь большую роль играют цифровые технологии, позволяющие оперативно фиксировать все необходимые изменения в его организме и прогнозировать результат лечения.

Однако внедрение этих технологий связано с рядом вызовов – от вопросов конфиденциальности и безопасности данных до этических аспектов использования ИИ в принятии медицинских решений.

В данной статье рассмотрена роль цифровых агентов в медицине, проанализированы новые возможности, которые они открывают для современного здравоохранения, а также вызовы, связанные с их внедрением. Определены перспективы развития цифровых агентов в медицине и их влияние на будущее здравоохранения. Обоснована важность сбалансированного подхода между дальнейшим развитием технологических инноваций и сохранением человеческого опыта в медицинской сфере.

Роль цифровых агентов в медицине и новые возможности

Цифровой агент представляет собой не человека, а технологию, которая в определенной степени имитирует его интеллект и имеет в настоящее время узкое практическое применение. В данной статье под термином «цифровой агент» понимаются все устройства с ИИ, обозначаемые как системы ИИ, нейронные сети, мобильные приложения, чат-боты и пр. Поскольку представители когнитивной науки и связанных с ней дисциплин, опираясь на воззрения математика А. Тьюринга и нейрофизиолога У. Мак-Каллока, рассматривают разум («mind») как вычислительное устройство, манипулирующее символами, цифровой агент должен иметь прежде всего вычислительные функции для обработки больших объемов структурированных данных. В настоящий момент цифровой агент может быть определен как представитель, помощник и ассистент человека. В то же время понятию «агент» присуща амбивалентность: с одной стороны агент – верный представитель, с другой стороны, агент автономен, самостоятелен и, следовательно, способен иметь собственные

цели. Свойство амбивалентности агента американский ученый в области ИИ М. Минский уподобил незаурядному уму раба и привел в связи с этим известный парадокс: «Если вы не будете давать ему (рабу) слишком многому учиться, вы снизите его полезность. Но если вы поможете ему стать умнее вас, вы не сможете гарантировать, что он не начнет строить для себя лучшие планы, чем для вас» [цит. по: Riecken, 1994, 25]. Такая альтернатива вполне может быть отнесена к цифровому агенту.

Использование термина «агент» при рассмотрении этических проблем обычно отсылает к проблематике «морального агента» и обширной дискуссии о том, могут ли системы ИИ быть таковыми. Определение сущности морального агента подразумевает антропоцентричность и понимание под «моральным агентом» любого агента, который надлежащим образом несет ответственность за действия [Parthemore, Whitby, 2013]. Существующие системы ИИ не могут являться моральными агентами, так как имеют базовую программу, определенную извне. «…ИИ не может быть моральным агентом с точки зрения моральной ответственности. Общество не может возлагать на ИИ ответственность за совершенные действия и принятые решения. Также в отношении ИИ не может быть ожидания моральной ответственности как осознания последствий за совершенные поступки» [Ларионов, Перова, 2023].

Несмотря на то, что в настоящее время искусственные интеллектуальные системы не являются моральными агентами, однако все более углубляющаяся связь человека и техники приводит к новым формам агентности, т.е. «по мере усложнения и усиления наших связей с машинами мы вновь и вновь сталкиваемся с вопросами, касающимися человекоподобных возможностей машин и машиноподобных свойств людей» [Сачмен, 2019, 19]. Это с очевидностью приводит к размышлению о том, возможно ли преодолеть ограничения машин путем кодирования в них все большего числа когнитивных способностей, присущих людям, и достигнут ли они человеческого уровня восприятия и понимания.

Поскольку современные технологии, включая ИИ и машинное обучение, становятся неотъемлемой частью диагностического процесса, цифровые агенты, к которым мы можем отнести системы поддержки принятия врачебного решения (СППВР), медицинские мобильные приложения и чат-боты, приобретают все большее значение в медицине. Они позволяют врачам совместно с пациентами проводить более точную и быструю диагностику, основываясь на анализе больших объемов медицинских данных. Использование алгоритмов машинного обучения позволяет выявлять скрытые закономерности и предсказывать возможные заболевания, даже на ранних стадиях развития.

Следует отметить, что одной из важных областей, где цифровые агенты приносят ощутимую пользу, является персонализированное лечение. Алгоритмы машинного обучения анализируют данные о пациенте, включая генетическую информацию, историю болезни и реакции на предыдущее лечение, для разработки индивидуальных планов лечения. Это дает возможность более эффективно бороться с различными заболеваниями, так как учитываются уникальные особенности каждого пациента. Например, при лечении онкологических заболеваний цифровые агенты могут предсказать, какой тип терапии

будет наиболее эффективен для конкретного вида рака, учитывая генетические мутации опухоли.

Цифровые агенты играют ключевую роль в постоянном мониторинге состояния здоровья пациентов. Носимые устройства, такие как умные часы и датчики, предоставляют непрерывные данные о физиологических показателях, активности и характеристиках сна. Эта информация может быть собрана в реальном времени и использована для создания динамичных картин состояния пациента. Цифровые агенты, осуществляющие постоянный мониторинг, который особенно важен для пациентов с хроническими заболеваниями, такими как сахарный диабет, сердечные заболевания или болезни дыхательной системы, позволяют на ранней стадии выявлять изменения в состоянии пациента и, соответственно, более эффективно на это реагировать, предупреждая возможные осложнения.

Внедрение персонализированного лечения также способствует снижению побочных эффектов, так как алгоритмы могут предсказывать, как организм отреагирует на определенные препараты. Очевидно, что сегодня врач не в состоянии запомнить все возможные взаимодействия лекарств, особенно в сложных случаях, когда пациент принимает десятки препаратов, и цифровой агент здесь может оказать большую помощь. Системы поддержки принятия врачебных решений, встраиваемые в медицинские картотеки, способны помочь врачам предотвратить фатальные результаты. СППВР занимается мониторингом и предупреждает врачей о состояниях пациентов, предписаниях и лечении, предоставляя основанные на научных данных клинические рекомендации [Committee on Patient Safety and Health Information Technology, 2012, 39]. Уже сейчас есть данные, свидетельствующие о том, что СППВР снижает количество ошибок [Moja, Kwag, 2014].

Все большую роль в медицине начинают играть чат-боты и медицинские мобильные приложения. Появляется много новых данных об эффективности их использования в здравоохранении. Было обнаружено, что чат-боты и диалоговые агенты могут быть полезны в области психического здоровья [Vaidyam et al., 2019] и при скрининге рака [Owen et al., 2018]. А осенью 2023 г. появилась информация, что ChatGPT (чат-бот с ИИ) помог установить редкий диагноз ребенку из США в то время, как 17 врачей в течение трех лет не могли определить болезнь и назначить правильное лечение [Joshu, 2023].

В качестве положительного примера использования цифровых агентов могут служить результаты исследования, проведенного с целью оценки способности ChatGPT предоставлять качественные ответы на вопросы пациентов [Ayers, Poliak, 2023]. Для исследования были случайно выбраны 195 вопросов пациентов с публичного форума в социальных сетях, на которые ответили как профессиональные врачи, так и чат-бот. Ответы были оценены командой медицинских работников по их качеству и проявлению эмпатии.

Ответы чат-ботов были признаны более предпочтительными и оценены выше по качеству и эмпатии, чем ответы врачей, т.е. показали преимущества чат-бота по сравнению с врачами. Авторы статьи пришли к выводу, что ChatGPT способен предоставлять качественные ответы на вопросы пациентов,

задаваемые на публичных форумах, которые впоследствии могут быть доработаны врачами и, в частности, помочь им в составлении ответов, что будет способствовать снижению уровня выгорания медицинских работников [Введенская, 2023].

По выполняемым функциям среди медицинских чат-ботов следует выделить бот под названием Omaolo, разработанный Финским институтом здравоохранения и социального обеспечения, который представляет собой онлайн-инструмент для оценки симптомов (электронный опросник) [Atique, Bautista, 2020]. Однако Omaolo не просто автоматизированный вопросник, это «электронный механизм медицинских знаний, который работает в соответствии с доказательной медицинской информацией…» и «объединяет симптомы, результаты измерений и медицинскую информацию, сообщаемую резидентом, с данными исследований» [THL, 2020, 10]. Omaolo может выполнить как оценку проблемы со здоровьем или выявить симптомы, так и направить пациента на получение лечения в конкретную медицинскую организацию. Omaolo также предлагает исследования качества жизни, обследования состояния полости рта и коучинг по вопросам здоровья. Интересно, что данный бот был в большей степени востребован для диагностики состояний, которые обычно считались интимными, таких как инфекции мочевыводящих путей и заболевания, передающиеся половым путем (ЗППП) [Pynnönen et al., 2020, 24].

Медицинские мобильные приложения для самодиагностики заболеваний могут быть полезны и в дерматологии. Если раньше у пациентов не было альтернатив в отношении постановки диагноза заболевания и назначения курса лечения, то теперь возможны различные варианты. Так, в настоящее время пациент обладает большей степенью автономности и сам решает, какой вариант диагностики и лечения для него наиболее приемлем.

Заинтересованы в медицинских приложениях также группы пациентов, которые стремятся избежать прямого контакта с врачом из-за возможной лишней информированности о своем заболевании или из-за недоверия врачам, больших очередей, плохого сервиса и по другим причинам. Доступность мобильных приложений для мгновенной диагностики дерматологических заболеваний, в том числе диагностики различных видов рака кожи, сделало возможным бесконтактное получение медицинских услуг. Созданное в России мобильное приложение «ProРодинки», например, позволяет по загруженным фотографиям подозрительных пятен на теле и родинок получить предварительное заключение о том, стоит ли продолжить самостоятельное наблюдение или необходимо срочно обратится к врачу, т.е. диагностика осуществляется, по сути, компьютерной программой без участия специалиста.

Новые возможности представляет также использование цифровых агентов в психиатрии. Медицинские мобильные приложения для людей с психическими проблемами, такие как Woebot, уменьшают стигматизацию, часто связанную с обращением пациентов за помощью к специалистам. Во многих странах мира из-за нехватки квалифицированных психологов и психиатров такие приложения становятся очень востребованными. Высокая стоимость традиционной терапии и табу в отношении психического здоровья также удерживают многих людей от использования традиционных вариантов психической

и психиатрической помощи, стимулируя использование доступных альтернатив [O'Sullivan, 2023].

Некоторые цифровые агенты сегодня проектируются для выполнения задач сопровождения и эмоциональной связи. Так, было создано анимированное приложение в виде кота (Sox), умеющее вести диалог и давать полезные советы, которое может стать незаменимым помощником для пожилых, одиноких людей. Оно, в частности, помогло скрасить одиночество мужчине, потерявшему жену [Паскуале, 2022, 105].

Несмотря на новые возможности, которые появляются у цифровых агентов в здравоохранении, во многих случаях и при различных нозологиях сохраняется неопределенность и сомнения как в постановке правильного диагноза, так и в подходящих методах лечения. Медицина – не точная наука, поэтому не всегда возможно рационально просчитать и предсказать наилучшие исходы для конкретного пациента. По справедливому замечанию Ф. Паскуале, современная медицина требует участия или, по крайней мере, понимания пациента, который должен соблюдать предложенный план лечения. Такие факторы значительно усложняют и обогащают отношения врача и пациента, в которых существенным оказывается собственно человеческая составляющая [Там же, 106].

Вызовы, связанные с использованием цифровых агентов в медицине

Наряду с приведенными в предыдущем разделе возможностями использования цифровых агентов и положительными примерами их применения в медицине, существуют и определенные вызовы. Выше отмечалась важность человеческой составляющей в медицине. Одной из областей, где особенно важен человеческий контакт, является психология. Рассмотренный опыт использования цифровых агентов в этой сфере, несмотря на преимущества, проанализированные в предыдущем разделе, в основном показывает сомнительные результаты. Например, национальная ассоциация расстройств пищевого поведения (NEDA) попыталась заменить всех сотрудников горячей линии психологической поддержки пациентов с расстройством пищевого поведения (РПП) чат-ботом, однако выяснилось, что он приносит больше вреда, чем пользы [Picchi, 2023].

Чат-бот Тесса был запущен в феврале 2022 г., созданный изначально как профилактический инструмент, предлагающий стратегии для предотвращения РПП. Впоследствии оказалось, что чат-бот давал сомнительные и даже вредные советы, которые могли ухудшить состояние пациентов. Например, он рекомендовал считать килокалории и снижать их до 1000 в день, измерять толщину жировых складок, что могло только усугубить РПП. Решение заменить «живых сотрудников» горячей линии чат-ботом имело негативные последствия из-за ограниченных возможностей бота в обеспечении эффективной и эмпатичной поддержки людей с РПП, что подчеркивает важность человеческого взаимодействия и участия в таких случаях [Введенская, 2023].

С этической точки зрения этот случай поднимает вопросы безопасности и эффективности чат-ботов в области психологической помощи. Недостаток регулирования и контроля за тем, как чат-бот представляет советы и оказывает поддержку, может привести к потенциальному вреду для пациентов. Важным, по моему мнению, является соответствие технологических инноваций этическим нормам, особенно в связи с заботой о пациентах с психологическими расстройствами. Имеется в виду соответствие инноваций принципализму, разработанному американскими философами Т. Бичампом и Дж. Чилдрессом. Принцип «не навреди» (non-maleficence) говорит о том, что врачи и другие специалисты из области здравоохранения должны стремиться к предотвращению вреда для пациентов. В случае с чат-ботами в области психологической помощи это означает, что разработчики должны гарантировать, что информация и советы, предоставляемые ботом, не могут причинить пациенту дополнительный психологический или эмоциональный вред. Это может включать в себя предоставление правильных и точных советов, а также способы обеспечения безопасного и эмпатичного взаимодействия с пациентами в кризисных ситуациях.

Принцип «делай благо» (beneficence) относится к обязательству специалистов здравоохранения действовать в интересах пациента и стремиться к их благополучию. В контексте использования чат-ботов для психологической помощи это означает, что боты должны быть разработаны с целью предоставления эффективной поддержки и советов, способных улучшить психологическое состояние пациентов. Это также может включать в себя предоставление ресурсов для самопомощи и своевременное направление пациентов к специалистам.

Принцип автономии означает уважение права пациента на принятие собственных решений относительно своего здоровья и лечения. В данном случае это означает, что чат-бот должен предоставить пациентам информацию для принятия осознанных решений относительно своего психического благополучия, например, предоставление пациентам выбора того, какие типы поддержки они хотели бы получать от чат-бота, а также предоставление информации о доступных вариантах лечения.

По мнению специалистов, чат-боты не подходят для лечения серьезных психических заболеваний, так как они сложны, имеют множество нюансов и уникальны для каждого человека, которого они затрагивают. В некоторых случаях предпочтительнее лечение с помощью лекарств, назначенных квалифицированным врачом. Как отмечает А. Бишт, важно избегать чрезмерного упрощения сложных проблем психического здоровья и поощрять пользователей обращаться за профессиональной помощью, когда это необходимо [O'Sullivan, 2023].

Учитывая конкретные примеры актуальности ограничения использования цифровых агентов в психологии и психиатрии, рассмотрим в целом «этические ловушки», без устранения которых, невозможно эффективное внедрение цифровых агентов в здравоохранение.

Проанализируем 4 вида «этических ловушек»:

1. Предвзятость. Известно, что система ИИ хороша настолько, насколько хороши данные, на которых она обучается. «Предвзятые данные» могут привести к негативным результатам для пациентов. В последние несколько лет были

публично продемонстрированы неудачи, когда алгоритмы фактически ухудшали уход за пациентами, вносили предвзятость, в частности, усложняя жизнь заболевшим афроамериканцам. Так, исследование, опубликованное в журнале Science в 2019 г., обнаружило расовую предвзятость в алгоритме, применявшемся для выявления пациентов со сложными медицинскими потребностями [Obermeyer, 2019]. В связи с тем, как указывается в исследовании, что на темнокожих больных с таким же, как у светлокожих уровнем потребностей тратится меньше денег, алгоритм ошибочно приходит к выводу, что темнокожие пациенты здоровее, чем такие же больные белые пациенты», – говорится в исследовании [Ibid.].

Авторы исследования подсчитали, что расовая предвзятость в алгоритме сократила количество чернокожих пациентов, нуждающихся в дополнительной помощи, более чем вдвое.

Группа ученых из США, Тайваня и Канады, опубликовавшая свою работу в журнале The Lancet Digital Health, обучила свой ИИ, используя сотни тысяч рентгеновских снимков, содержащих информацию о расе пациента, определять этническую принадлежность пациентов [Gichoya, Banerjee, 2022]. Нейросеть угадывала расу человека (белый, темнокожий, азиат) с 98%-й точностью, даже когда сканы брали у людей одного и того же возраста, пола и комплекции. «Мы показываем, что стандартные модели глубокого обучения ИИ могут определять расу по медицинским снимкам с высокой точностью в нескольких модальностях визуализации. Это поднимает тревожные вопросы о роли ИИ в медицинской диагностике и лечении: могут ли модели непреднамеренно проявлять расовую предвзятость при изучении подобных изображений?» [Ibid.], – говорится в статье. Это исследование еще раз подтверждает, что цифровые агенты могут отражать различные предубеждения людей, будь то расизм, сексизм и др. Религиозные, психические, культурные и образовательные различия создателей алгоритмов могут повлиять на их этические представления и привести к определенным искажениям в данных. Искажения же в данных, на которых обучается ИИ, приводят к искаженным результатам, что делает их потенциально опасными для общества. Такая предвзятость является серьезной этической проблемой в области ИИ, которая может углубить всевозможное неравенство, дискриминацию людей в сфере здравоохранения и повлиять на решения миллионов людей в отношении лечения, что может оставить их без необходимой помощи.

Таким образом, становится очевидным вывод о необходимости выявления предвзятости данных, прежде чем они станут основой системы, применяющейся в здравоохранении.

2. Галлюцинации и вредная информация. Цифровые агенты могут галлюцинировать, т.е. выдумывать несуществующую информацию. Очевидно, что цифровой агент, проявляющий творческий подход в контексте здравоохранения, вызывает проблемы. Понятно, что выдуманная или ложная информация, особенно когда она используется при принятии клинических решений, может нанести вред пациенту. В связи с этим необходимо понять, когда, где и почему эти модели действуют определенным образом, и провести оперативную разработку, чтобы они не ошибались по причине собственного творчества.

Даже если информация, которую цифровой агент предлагает пользователю, не является продуктом галлюцинации, следует быть уверенным, что она служит своему прямому назначению. Выше мы уже рассмотрели пример потенциального вреда, который может нанести чат-бот Тесса людям с РПП.

При использовании цифровых агентов как врачами, так и пациентами всегда следует иметь в виду, что методам ИИ, как правило, не хватает «здравого смысла», что делает их неспособными выявлять простые ошибки в данных или решениях, которые были бы очевидны для человека [McDermott, DeLaurentis, 2020].

Поглощение цифровыми агентами миллионов страниц медицинской информации не тождественно ее осмыслению и разумному использованию, поскольку сами данные, синтаксис (форма) и статистика еще не порождают семантического понимания.

Когда врачи наблюдают за пациентом с конкретными симптомами, они оценивают субъективную вероятность диагноза, при этом на практике диагнозы ставятся более сложным образом, который врачи редко могут объяснить логически [Banerjee et al., 2009]. В отличие от искусственных систем, опытные врачи признают тот факт, что диагнозы и прогнозы всегда отмечены разной степенью неопределенности. Они осознают, что некоторые диагнозы могут оказаться неверными и некоторые методы лечения могут не привести к ожидаемому излечению.

Таким образом, медицинский диагноз и принятие решений требуют «благоразумия», в соответствии с определением Аристотеля, или «практической мудрости», которые относятся к гибкой способности к интерпретации, позволяющей врачу определить лучший образ действий, когда знания зависят от обстоятельств [Montgomery, 2006]. Квалифицированные врачи, будучи разумными субъектами, способны работать в сложной триаде медицинской практики, этических стандартов и научных знаний. Врачам часто нужна мудрость, а не интеллект, и мы очень далеки от науки об искусственной мудрости [Powell, 2019, 2].

3. Конфиденциальность и согласие субъекта данных. С широким использованием цифровых агентов в медицине возникают серьезные вопросы о конфиденциальности и безопасности медицинских данных. ИИ в здравоохранении и любой другой отрасли требует больших данных. Откуда берутся данные, как они передаются – остается неизвестным для большинства пациентов. Они не знают, используется ли их личная информация и каким образом. Защищены ли данные от посторонних и злоумышленников, если они используются системой ИИ? Могут ли пользователи доверять компании, использующей инструменты ИИ, что она сохранит их конфиденциальность?

Поскольку цифровые агенты все больше интегрируются в здравоохранение, требования сохранения конфиденциальности информации о пациентах становятся более строгими. Пациенты имеют право хранить свою медицинскую информацию в тайне, но имеют ли они право знать, используется ли ИИ в их лечении или их данные используются для обучения ИИ?

Очевидно, что сбор и обработка больших объемов «чувствительной» информации о здоровье пациентов требует высоких стандартов защиты, чтобы

предотвратить несанкционированный доступ и утечку данных. Одним из решений этой проблемы является разработка и внедрение строгих мер безопасности, таких как криптографическое шифрование данных, двухфакторная аутентификация, обеспечивающие защиту от хакерских атак. Кроме того, необходимо разработать законодательство, регулирующее сбор, хранение и использование медицинских данных, чтобы обеспечить их адекватную защиту и соблюдение прав пациентов.

4. Прозрачность и объяснимость. Термин «черный ящик» часто употребляется, когда анализируется способ, согласно которому цифровой агент пришел к тому или иному решению. В нейронной сети могут возникать ложные корреляционные зависимости и ошибки, и если система приняла неправильное решение, то невозможно найти ответ, почему. Ответ, таким образом, может оказаться в ловушке этого черного ящика, поскольку нет возможности проследить процесс принятия решений системой.

Однако трудно переоценить важность объяснимости решений цифровых агентов, а также постоянного присутствия человека для контроля и исправления любых ошибок, возникающих в результате использования системы ИИ. Когда люди выпадают из этого цикла, результаты могут быть катастрофическими.

Таким образом, объяснимость – это сложное, но основополагающее требование для всех технологий здравоохранения с поддержкой ИИ [Pawar, O'Shea, 2020]. Согласно этому принципу, разработчики алгоритмов должны быть способны объяснить общее обоснование и методологию решения ИИ, выборку данных, используемых для обучения модели, а также решения и действия по управлению данными, связанные с ними. Этот принцип особенно актуален в медицинских учреждениях, где у пациентов есть законная причина знать, как ИИ обнаружил конкретное заболевание и на каких факторах была основана рекомендация, относящаяся к здоровью [Sand, Durán, 2022].

Кроме рассмотренных выше «этических ловушек», которые необходимо преодолеть для эффективного использования цифровых агентов в медицине, большую обеспокоенность вызывает все большее делегирование ИИ принятия решений о диагнозе и лечении, ограничивающее автономию врачей в пользу предварительно заданных алгоритмов. Активное внедрение в клиническую практику цифровых агентов чревато также потерей компетенций и навыков врачей и в целом переоценкой роли клинического мышления.

Рассматривая ограничение автономии врачей, нельзя не обратить внимания на автономию пациентов. Так, М. Заляускайте предполагает, что такие технологии, как мобильные приложения, которые используются пациентами для самоконтроля (сбора любых форм медицинских данных), могут повысить автономию больных и в лучшем случае перевести отношения между врачом и пациентом в формат обслуживания, где обе стороны имеют сбалансированное распределение прав и обязанностей [Žaliauskaitė, 2020]. Однако такое распределение прав и обязанностей в отношениях между врачом и пациентом, которые обычно характеризуются уязвимостью пациента по отношению к врачу, мне кажется сомнительным. При рассмотрении типа автономии, которую смог бы поддерживать ИИ, неясной представляется возможность алгоритма учитывать предпочтения разных людей (в отношении лечения) [McDougall, 2018].

Это может привести к появлению новой формы «цифрового патернализма», при которой ИИ принимает решения от имени пациентов и врачей с той разницей, что патерналистские отношения будут строиться по отношению к алгоритму, а не к врачу.

В этой связи возникают следующие этические вопросы: как обеспечить, чтобы пациенты сохраняли контроль над своими медицинскими решениями при использовании цифровых агентов? Как избежать ситуаций, когда технологии начинают доминировать над личными правами и предпочтениями пациентов?

Феномен «цифрового патернализма» влечет за собой и утрату доверия к квалифицированным врачам. Поскольку одной из характеристик цифровых агентов является отсутствие человеческого общения, что является противоположностью традиционному личному взаимодействию пациента с врачом, они могут усилить недоверие к медицинским услугам, предоставляемым людьми.

Дальнейшее распространение цифровых агентов в медицине может привести не только к существенному изменению во взаимоотношениях врачей и пациентов, но и к социальному расслоению, к разделению общества на тех, кто пользуется человеческими связями, и тех, кто вынужден прибегать к помощи цифровых агентов.

В итоге все более актуальной становится проблема справедливого доступа к медицинским услугам. Если цифровые агенты станут основным источником заботы о здоровье, это может создать барьеры как для людей с ограниченным доступом к технологиям, так и для людей, предпочитающих традиционные подходы к оказанию медицинской помощи.

Кроме анализа и предотвращения этических коллизий, возникающих в процессе использования цифровых агентов в медицине, важно обратить внимание на этичность самой технологии ИИ. Именно здесь кроется основная сложность, поскольку этические нормы трудно формализовать и заложить в ИИ. В этике существуют две основные доктрины: утилитаризм и абсолютизм. Возникает вопрос: какую этическую доктрину из них надо будет принять при программировании цифровых агентов – абсолютистскую или утилитарную? Если возникнет дилемма: можно спасти жизни нескольких человек ценой жизни одного, но того, кто в принципе не виноват в возникновении самой критической ситуации. Абсолютистский подход, конечно, однозначно запрещает приносить в жертву какого-либо одного случайно попавшего в ситуацию человека ради спасения жизней нескольких людей. А утилитарный подход при каких-то обстоятельствах это в принципе допускает [Разин, 2019]. Таким образом, выбор подхода, который будет заложен в конкретную систему ИИ, достаточно сложен для разработчиков алгоритмов, нуждается в общественном обсуждении, а также требует этической рефлексии морального философа, который способен более точно определить этические рамки и правила, которые будут определять функционирование системы.

Рассмотренные выше этические проблемы свидетельствуют о необходимости тщательного обсуждения и разработки правил и нормативов для внедрения цифровых агентов в медицину с учетом максимального уважения прав пациентов, человеческих отношений и этических ценностей. В настоящее время в мире разрабатывается множество этических кодексов ИИ, регулирующих возникающие или

потенциально возможные этические проблемы. Диапазон этих проблем достаточно широк: от нарушения конфиденциальности и безопасности до явных угроз человеку, поэтому сегодня предпринимаются многочисленные попытки их стандартизации. Этические вопросы, касающиеся ИИ, стоят перед всем мировым сообществом, поэтому необходимо разработать документальную, нормативную основу, которой смогут следовать все страны, чтобы на ее основе стало возможным сформулировать уточняющие стандарты или рекомендации, учитывающие собственные ценности, культурные традиции, моральные нормы различных стран.

Заключение

В свете перспективных возможностей цифровых агентов в медицине необходимо уделять внимание не только технологическим достижениям, но и рассматривать вызовы, к которым они приводят, концентрируясь на этических аспектах их использования. Защита конфиденциальности данных, устранение предвзятости и галлюцинаций алгоритмов, а также обеспечение прозрачности работы цифровых агентов и соблюдение норм этики врачебной практики остаются важными составляющими внедрения цифровых технологий в медицину. Баланс между технологическими инновациями и человеческим опытом представляет собой неотъемлемую часть успешного развития здравоохранения. Врачи и медицинский персонал продолжают играть ключевую роль в процессе лечения, при этом цифровые агенты должны становиться инструментами, усиливающими их способности, а не заменяющими их. В итоге будущее здравоохранения формируется во взаимодействии человека и технологии. Сбалансированный подход, ориентированный на пациента, для которого важен человеческий фактор в медицинской практике, станет краеугольным камнем для успешной реализации перспектив цифровых агентов в медицине. Внимание к этим аспектам позволит создать здравоохранение, которое будет не только эффективно и доступно, но и гуманно по своей сути.

Digital Agents in Medicine: New Opportunities and Challenges

Elena V. Vvedenskaya

Pirogov Russian National Research Medical University. 1 Ostrovitianov Str., Moscow, 117997, Russian Federation.

ORCID: 0000-0003-1100-0033
e-mail: vvedenskaya.elena@gmail.com

The subject of this article is the new opportunities and challenges that the widespread use of digital agents in medicine brings. ‘Digital agents’ refers to all AI devices, neural networks, mobile applications, chatbots used for diagnostics and patient monitoring. Along with the advantages of their introduction, there are a number of risks provoked by the loss of the human factor and lack of regulation of new technologies. The article discusses

the “ethical pitfalls”, without the elimination of which the effective implementation of digital agents in healthcare is impossible. These include: bias, hallucinations and harmful information, data subject privacy and consent, as well as transparency and explainability. It has been observed that there is an increasing delegation of diagnosis and treatment decisions to AI, limiting the autonomy of physicians in favor of pre-defined algorithms. This phenomenon of “digital paternalism” entails a loss of trust in qualified physicians. In addition, the article analyzes the problem of social stratification caused by both limiting access to modern technologies for certain segments of the population and creating barriers for people who prefer traditional approaches to medical care. It concludes that it is necessary to thoroughly discuss and develop rules and regulations for the introduction of digital agents in medicine, taking into account the maximum respect for patients’ rights, preservation of human relations “doctor–patient” and ethical values.

Keywords: artificial intelligence, digital agents, medicine, ChatGPT, ethics, bias, confidentiality, explicability, prudence

Литература / References

Введенская Е.В. Трансформация взаимоотношений врача и пациента: от биоэтики к роботике // Человек. 2023. Т. 34. № 6. С. 65–83.

Vvedenskaya, E.V. “Transformaciya vzaimootnoshenij vracha i pacienta: ot bioetiki k roboetike” [Transformation of the Physician-patient Relationship: from Bioethics to Roboethics], *Chelovek*, 2023, Vol. 34, No. 6, pp. 65–83. (In Russian)

Ларионов И.Ю., Перова Н.В. «Машина бога» Дж. Савулеску как моральный агент и проблема ответственности // Человек. 2023. Т. 34. № 3. С. 24–40.

Larionov, I., Perova N. “‘Mashina boga’ Dzh. Cavulesku kak moral’nyj agent i problema otvetstvennosti” [J. Savulescu’s [“The God Machine” as a Moral Agent and the Problem of Responsibility], *Chelovek*, 2023, Vol. 34, No. 3, pp. 24–40. (In Russian)

Паскуале Ф. Новые законы робототехники: апология человеческих знаний в эпоху искусственного интеллекта / Пер. с англ. А. Королева. М: Издат. дом «Дело» РАНХиГС, 2022.

Paskuale, F. *Novye zakony robototekhniki: apologiya chelovecheskih znanij v epohu iskusstvennogo intellekta* [New Laws of Robotics: an Apology for Human Knowledge in the Era of Artificial Intelligence], transl. by A. Korolev. Moscow: Izdatel’skij dom «Delo» RANHiGS Publ., 2022. (In Russian)

Разин А.В. Этика искусственного интеллекта // Философия и общество. 2019. № 1 (90). URL: https://cyberleninka.ru/article/n/etika-iskusstvennogo-intellekta (дата обращения: 17.03.2024).

Razin, A.V. “Etika iskusstvennogo intellekta” [Ethics of Artificial Intelligence], *Filosofiya i obshchestvo,* 2019, No. 1 (90) [https://cyberleninka.ru/article/n/etika-iskusstvennogo-intellekta, accessed on 17.03.2024]. (In Russian)

Сачмен Л. Реконфигурации отношений человек – машина: планы и ситуативные действия / Пер. с англ. А.С. Максимовой. М.: Элементарные формы, 2019.

Sachmen, L. *Rekonfiguracii otnoshenij chelovek – mashina: plany i situativnye dejstviya* [Human – Machine Reconfigurations: Plans and Situated Actions], transl. by A.S. Maksimova. Moscow: Elementarnye formy Publ., 2019. (In Russian)

Atique, S., Bautista, J.R. et al. “A Nursing Informatics Response to COVID-19: Perspectives From Five Regions of the World”, *Journal of Advanced Nursing*, 2020, No. 76 (10), pp. 2462–2468.

Banerjee, A., Jadhav, S.L., Bhawalkar, J.S. “Probability, Clinical Decision Making and Hypothesis Testing”, *Industrial Psychiatry Journal*, 2009, No. 18 (1), pp. 64–69.

Committee on Patient Safety and Health Information Technology Board on Health Care Services, Health IT and Patient Safety: Building Safer Systems for Better Care. Washington, DC: The National Academies Press, 2012.

Gichoya, J.W., Banerjee, I., Bhimireddy, A.R. "AI Recognition of Patient Race in Medical Imaging: a Modelling Study", *The Lancet Digital Health*, 2022, Vol. 4 [https://doi.org/10.1016/S2589-7500(22)00063-2, accessed on 22.01.2024].

Joshu, E. Toddler Whose Symptoms Puzzled 17 Doctors for Three YEARS is Finally Diagnosed With Rare Condition… by ChatGPT, The Daily Mail, 2023, Sept. 12 [https://www.dailymail.co.uk/health/article-12509111/ChatGPT-diagnosis-rare-condition.html, accessed on 22.01.2024].

McDermott, T., DeLaurentis, D. et al. "AI4SE and SE4AI: A Research Roadmap", *INSIGHT*, 2020, No. 23 (1), pp. 8–14.

McDougall, RJ. "Computer Knows Best? The Need for Value-flexibility in Medical AI", *Journal of Medical Ethics*, 2018, No. 45 (3), p. 156.

Moja, L., Kwag, K.H. et al. "Effectiveness of Computerized Decision Support Systems Linked to Electronic Health Records: A Systematic Review and Meta-Analysis", *American Journal of Public Health*, 2014, pp. 12–22.

Montgomery, K. *How Doctors Think: Clinical Judgment and the Practice of Medicine*. Oxford: Oxford UP, 2006.

Obermeyer, Z. et al. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations", *Science*, 2019, No. 366, pp. 447–453.

O'Sullivan, I. *Why AI Therapy Chatbots Are the Ultimate Ethical Dilemma* [https://tech.co/news/ai-therapy-chatbots-ethical-risks, accessed on 22.01.2024].

Owens, O.L, Felder, T. et al. "Evaluation of a Computer-based Decision Aid for Promoting Informed Prostate Cancer Screening Decisions Among African American Men: iDecide", *American Journal of Health Promotion*, 2019, No. 33 (2), pp. 267–278.

Parthemore, J., Whitby, B. "What Makes any Agent a Moral Agent? Reflections on Machine Consciousness and Moral Agency", *International Journal of Machine Consciousness*, 2013, No. 05 (02), pp. 1–21.

Pawar, U., O'Shea, D., Rea, S. O'Reilly, R. "Explainable AI in Healthcare", *Proceedings of the 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. Dublin, Ireland, 15–19 June 2020. Piscataway, NJ: IEEE, 2020, pp. 1–12.

Picchi, A. "Eating Disorder Helpline Shuts Down AI Chatbot that Gave Bad Advice", *CBS Interactive Inc*. 2023, June 1, [https://www.cbsnews.com/news/eating-disorder-helpline-chatbot-disabled/, accessed on 22.01.2024].

Powell, J. "Trust Me, I'm a Chatbot: How Artificial Intelligence in Health Care Fails the Turing Test", *Journal of Medical Internet Research*, 2019, No. 21 (10): e16222. [https://pubmed.ncbi.nlm.nih.gov/31661083/, accessed on 22.01.2024].

Pynnönen, T., Rantala, K., Räsänen, R. *Kokemuksia Omaolo-palvelusta (Experiences with the Omaolo Service). Bachelor's thesis*. Tampere: Tampere University of Applied Sciences, 2020.

Riecken, D. "A Conversation With Marvin Minsky about Agents", *Communications of the ACM*, 1994, Vol. 37, No. 7, pp. 23–29.

Sand, M., Durán, J.M., Jongsma, K.R. "Responsibility Beyond Design: Physicians' Requirements for Ethical Medical AI, *Bioethics*, 2022, No. 36, pp. 162–169.

THL – Finnish institute for health and welfare. 2020. Omaolo – Instructions for Use [https://www.omaolo.fi/, accessed on 22.01.2024].

Vaidyam, A.N. et al. "Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape", *Can J Psychiatry*, 2019, No. 64 (7), pp. 456–464.

Žaliauskaitė, M. "Role of Ruler or Intruder? Patient's Right to Autonomy in the Age of Innovation and Technologies", *AI & Society: Knowledge, Culture and Communication*, 2021, Vol. 36, pp. 573–583.