

В редакцию журнала поступило письмо читателя, в котором поднимаются актуальные вопросы, связанные с этическими аспектами использования искусственного интеллекта (ИИ) в медицинской практике.

«С каждым годом технологии ИИ все активнее внедряются в лучевую и функциональную диагностику, анализ данных компьютерной томографии, спирометрии и гистологических препаратов. Однако беспокоит не столько качество самой работы ИИ, сколько человеческий фактор в его использовании. Со временем, при регулярно высокой точности алгоритмов, у врача неизбежно может сформироваться так называемая “гипердоверенность” (*automation bias*). Возникает риск того, что специалист перестанет перепроверять заключения, сформированные нейросетью, и будет воспринимать их как истину в последней инстанции. Диагностика может превратиться в формальность: получил результат от ИИ — подписал. В связи с этим возникает этический вопрос: где находится грань дозволенного, за которой заканчивается разумное использование помощника и начинается подмена клинического мышления машиной? А также стоит ли вообще отдавать такие критические этапы, как первичная диагностика и постановка окончательного диагноза, в руки ИИ, если это потенциально ведет к “расслаблению” врача и утрате его настороженности? Как сохранить баланс, чтобы ИИ оставался помощником, а не подменял собой ответственность специалиста за жизнь пациента?».

На вопросы читателя отвечает руководитель направления искусственного интеллекта компании «ГЭОТАР-Цифра», руководитель проекта «Консультант врача» *Игорь Маратович Еникеев*.

Проблему чрезмерного доверия к системам искусственного интеллекта (ИИ) следует рассматривать как один из новых факторов риска в современной клинической практике. При этом проблема не в самой технологии и не в противопоставлении врача и алгоритма. Риск возникает тогда, когда результат работы алгоритма начинает определять сам способ принятия клинического решения.

В психологии деятельности человека этот феномен известен как *automation bias* — склонность принимать решение, предложенное автоматизированной системой, и недостаточно активно искать данные, которые могли бы этому решению противоречить [1]. В медицине подобная ошибка имеет особое значение. Заключение ИИ может быть сформулировано убедительно, сопровождаться количественной оценкой вероятности и появляться на экране раньше, чем врач успел составить собственное мнение. В этих условиях предложенная алгоритмом версия становится исходной гипотезой, от которой врачу психологически труднее отказаться.

Однако считать развитие такого гипердоверия неизбежным было бы неправильно. Риск зависит от того, как ИИ включен в клинический процесс, какую именно задачу ему поручили, каким образом врачу предъявляется результат и насколько специалист знает ограничения используемой системы. Один и тот же алгоритм способен уменьшить число пропущенных патологических изменений и одновременно стать источником дополнительной ошибки, если его неверная подсказка принимается без критической проверки.

Эта проблема уже получила экспериментальное подтверждение. В исследовании, посвященном выявлению рака легкого по результатам рентгенографии органов грудной клетки, врачи-рентгенологи интерпретировали рентгенограммы с дополнительной

подсказкой, представленной как результат работы ИИ. Когда такая подсказка была намеренно ошибочной, специалисты значительно чаще принимали неправильное решение в тех случаях, которые ранее интерпретировали правильно [2]. Таким образом, ошибочная рекомендация автоматизированной системы могла изменить первоначально верное решение врача. В другом многоцентровом проспективном рандомизированном исследовании с участием 220 врачей было показано, что тип объяснения, сопровождающего рекомендацию ИИ, может влиять на диагностическую точность и склонность врача следовать этой рекомендации. При этом сами участники не всегда осознавали такое влияние. Это указывает на то, что риск чрезмерного доверия определяется не только точностью алгоритма, но и способом представления его результата врачу [3].

Помимо непосредственного влияния подсказки ИИ на принимаемое решение, возникает другой вопрос: не меняется ли со временем способность специалиста самостоятельно выполнять диагностическую задачу без автоматизированной поддержки? По данным многоцентрового наблюдательного исследования в области эндоскопии было показано, что после внедрения системы компьютерного выявления полипов показатель выявления аденом при колоноскопиях, выполнявшихся без использования ИИ, снизился с 28,4 до 22,4 %. Наблюдательный характер исследования не позволяет утверждать, что причиной этого снижения стало использование ИИ. Однако полученные данные указывают на необходимость учитывать и специально изучать возможное влияние длительной автоматизированной поддержки на самостоятельную работу специалиста [4].

Где в таком случае проходит допустимая граница? Эту границу нельзя определять лишь тем, участвует ли

ИИ на этапе первичной диагностики или при установлении окончательного диагноза. В современной медицине многие диагностические операции уже частично автоматизированы. Алгоритм может измерять объемы и плотность легочной ткани, выявлять и маркировать очаговые образования, количественно оценивать эмфизему, анализировать параметры функции внешнего дыхания, сортировать исследования по срочности. Однако передача подобных операций компьютерной системе сама по себе не означает отказ врача от клинического мышления.

Принципиальное различие существует между оценкой отдельного диагностического признака и установлением диагноза как клинического решения. Диагноз не является производной одного изображения, показателя спирометрии или морфологического препарата. Он формируется при сопоставлении жалоб и анамнеза, данных физикального обследования, лабораторных и инструментальных исследований, динамики заболевания, вероятности альтернативных состояний и возможных последствий принятого решения для конкретного пациента. Даже очень точная система ИИ действует в пределах тех данных и той задачи, которые ей были заданы. Поэтому результат ее работы следует рассматривать как диагностическую информацию или диагностическую гипотезу, но не как самостоятельный эквивалент окончательного клинического диагноза.

Наиболее оправданной представляется модель, при которой ИИ расширяет возможности специалиста: обращает внимание на признак, который мог быть пропущен, выполняет количественный анализ, сопоставляет большой объем информации, помогает определить приоритет исследования или предлагает возможный дифференциально-диагностический ряд. Для отдельных строго ограниченных и хорошо валидированных задач допустима и более высокая степень автономности. Однако результаты, полученные для одной задачи, нельзя автоматически переносить на другую. Чем выше цена диагностической ошибки и чем больше решение зависит от индивидуального клинического контекста, тем более значимой становится независимая врачебная проверка.

При этом само участие врача в процессе принятия решения еще не гарантирует безопасности. Если специалисту фактически оставлена роль человека, который должен только подтвердить уже сформированное системой заключение, такой контроль быстро становится формальным. Содержательный врачебный контроль предполагает знание области применения алгоритма, его типичных ошибок и ограничений, возможность не согласиться с результатом и наличие достаточной информации для такого решения. В диагностически сложных ситуациях полезен принцип независимой первичной оценки: врач формирует собственное заключение, затем знакомится с результатом ИИ и разбирает расхождения. В задачах срочной сортировки исследований последовательность может быть иной, поскольку смысл алгоритма как раз состоит в раннем привлечении внимания к потенциально опасной находке. Но и в этом случае тревожный

сигнал ИИ не заменяет последующей медицинской оценки.

Этот подход постепенно становится частью международной системы регулирования. Всемирная организация здравоохранения относит сохранение автономии человека к основным принципам применения ИИ в здравоохранении [5]. В Регламенте Европейского союза об ИИ для систем высокого риска отдельно предусмотрена необходимость учитывать склонность человека к чрезмерному доверию результатам автоматизированной системы и сохранять возможность отклонить или изменить ее рекомендацию [6]. В современной лучевой диагностике все большее значение придется не только испытанию системы до ее внедрения, но и контролю за ее работой после внедрения в клиническую практику.

Такая оценка необходима потому, что точность ИИ не является неизменной величиной. Результаты могут отличаться при использовании другого оборудования, иной структуры пациентов, изменении протоколов исследования или обновлении программного обеспечения. Медицинской организации следует проводить локальную проверку системы, отслеживать ее работу после внедрения, анализировать расхождения между врачом и алгоритмом, контролировать ложноположительные и ложноотрицательные результаты. Для технологий, которые способны влиять на профессиональный навык, целесообразно сохранять и периодическую оценку работы врача без автоматизированной поддержки. Особенно важен этот принцип при подготовке молодых специалистов. Сначала необходимо научиться видеть патологический признак и строить диагностическое рассуждение самостоятельно; машинная подсказка не должна заменять этап формирования этого навыка.

Особого обсуждения требует проблема ответственности. Нельзя считать правильной ситуацию, при которой разработчик отвечает только за техническое функционирование программы, медицинская организация — только за ее внедрение, а все последствия применения автоматически возлагаются на врача, которому был показан результат ИИ. Безопасность является совместной задачей разработчика, регулятора, медицинской организации и медицинского работника. Вместе с тем клиническое решение, которое врач принимает и подтверждает своей подписью, не может быть фактически передано алгоритму только потому, что последний демонстрирует высокую статистическую точность. Врач не должен превращаться в формальный механизм подтверждения машинного заключения.

Существует и обратная сторона проблемы. Недоверие к ИИ само по себе не является показателем высокого профессионализма. Если клинически валидированная система помогает обнаружить изменения, которые человек может пропустить, отказ от такой помощи также способен ухудшить качество диагностики. Поэтому речь должна идти не о максимальном доверии или недоверии, а о доверии, соразмерном реальным возможностям и ограничениям конкретной технологии.

ИИ будет все глубже входить в лучевую диагностику, функциональные исследования, морфологию и системы поддержки принятия клинических решений. Поэтому вопрос состоит не в отказе от ИИ как такового, а в определении условий его безопасного применения. Задача медицинского сообщества состоит в том, чтобы определить правила его безопасного применения. ИИ способен взять на себя часть рутинной аналитической работы, но он не должен устранять

необходимость самостоятельного клинического суждения. Сохранение клинического мышления обеспечивается не отказом от технологии, а организацией работы, при которой врач способен сформулировать собственную позицию, сопоставить ее с результатом алгоритма и при наличии оснований отвергнуть его. Способность обоснованно не согласиться с ИИ остается одним из существенных признаков профессиональной ответственности врача.

Литература / References

1. Goddard K., Roudsari A., Wyatt J.C. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *JAMIA*. 2012; 19 (1): 121–127. DOI: 10.1136/amiajnl-2011-000089.
2. Bernstein M.H., Atalay M.K., Dibble E.H. et al. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. *Eur. Radiol.* 2023; 33 (11): 8263–8269. DOI: 10.1007/s00330-023-09747-1.
3. Prinster D., Mahmood A., Saria S. et al. Care to explain? AI explanation types differentially impact chest radiograph diagnostic performance and physician trust in AI. *Radiology*. 2024; 313 (2): e233261. DOI: 10.1148/radiol.233261.
4. Budzyń K., Romańczyk M., Kitala D. et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol. Hepatol.* 2025; 10 (10): 896–903. DOI: 10.1016/S2468-1253(25)00133-5.
5. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. Available at: <https://www.who.int/publications/i/item/9789240029200>
6. Regulation (EU) 2024/1689 of the European Parliament and of the council of 13 June 2024 laying down harmonised rules on artificial intelligence (artificial intelligence act). *OJEU*. 2024; L 2024/1689: Art. 14. Available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

Информация об авторе / Authors Information

Еникеев Игорь Маратович — врач-кардиолог; руководитель направления искусственного интеллекта компании «ГЭОТАР-Цифра», руководитель проекта «Консультант врача»; организатор здравоохранения; тел.: (916) 439-54-93; e-mail: enikeev@upzdrav.ru (SPIN-код: 5626-4892; ORCID: <https://orcid.org/0009-0004-7866-068X>)

Igor M. Enikeev, Cardiologist; Head of Artificial Intelligence at “GEOTAR-Tsifra”, Head of the “Physician Consultant” project; Healthcare Administrator; tel.: (916) 439-54-93; e-mail: enikeev@upzdrav.ru (SPIN-code: 5626-4892; ORCID: <https://orcid.org/0009-0004-7866-068X>)