

Научная статья



DOI: 10.55959/MSU2073-2643-21-2025-4-83-101

ДОВЕРИЕ К ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ В ЗДРАВООХРАНЕНИИ И ПУТИ ЕГО УКРЕПЛЕНИЯ

Еникеев И.М.

ООО «Консультант студента», Москва, Российская Федерация
enikeevigor99@gmail.com

Солнцев В.И.

НП «ИВМ Консалтинг Групп», Москва, Российская Федерация
vi.solntsev@gmail.com

Еникеев М.А.

ГБУЗ РК «Керченская больница № 1 имени Н.И. Пирогова», Керчь, Российская Федерация
marat_misha@mail.ru

Аннотация. В статье рассматриваются вопросы доверия к системам искусственного интеллекта (ИИ) в здравоохранении, которые становятся ключевым барьером на пути широкого внедрения новых технологий. Актуальность работы обусловлена быстрым ростом числа медицинских изделий на основе ИИ как в России, так и за рубежом, и одновременным дефицитом доверия со стороны профессионального сообщества и пациентов. В качестве методов исследования использован анализ современной научной литературы, международных стандартов, регуляторных практик, а также обзор эмпирических данных, полученных в зарубежных и отечественных исследованиях. Основное внимание уделяется восприятию ИИ пациентами и медицинскими специалистами, а также причинам скептицизма, связанным с непрозрачностью алгоритмов, риском ошибок и дискриминации. В статье обсуждаются стратегические направления по укреплению доверия: обеспечение прозрачности и объяснимости решений ИИ, проведение независимых испытаний и постмаркетингового мониторинга, вовлечение профессионального сообщества и пациентов в процесс внедрения технологий, а также развитие механизмов управления качеством на основе международных стандартов. Сделан вывод о необходимости формирования «калиброванного» доверия на основе эмпирически подтвержденной пользы и справедливости применения ИИ

в клинической практике. Перечислены задачи для дальнейших исследований в российских условиях.

Ключевые слова: искусственный интеллект, доверие, здравоохранение, прозрачность, объяснимость, медицинские технологии, регулирование.

Для цитирования: Еникеев И.М., Солнцева В.И., Еникеев М.А. Доверие к искусственному интеллекту в здравоохранении и пути его укрепления // Вестник Московского университета. Серия 21. Управление (государство и общество). 2025. Т. 22. № 4. С. 83–101.

Дата поступления в редакцию: 15.09.2025

TRUST IN ARTIFICIAL INTELLIGENCE SYSTEMS IN HEALTHCARE: BARRIERS AND STRATEGIES FOR IMPROVEMENT

Enikeev I.M.

LLC “Student Consultant”, Moscow, Russian Federation
enikeevigor99@gmail.com

Solntsev V.I.

NP “IVM Consulting Group”, Moscow, Russian Federation
vi.solntsev@gmail.com

Enikeev M.A.

State Budgetary Healthcare Institution of the Republic of Crimea “Kerch Hospital No. 1 named after N.I. Pirogov”, Kerch, Russian Federation
marat_misha@mail.ru

Abstract. This article addresses trust in artificial intelligence (AI) systems in healthcare, which remains a critical barrier to widespread adoption of advanced technologies. The relevance of this topic is highlighted by the rapid increase in the number of AI-based medical devices, both in Russia and worldwide, and the concurrent lack of trust from healthcare professionals and patients. The study uses a comprehensive analysis of contemporary scientific literature, international standards, regulatory practices, and a review of empirical data from global and Russian studies. The article focuses on the attitudes of patients and healthcare providers towards AI, examining the roots of skepticism related to algorithmic opacity, risks of error, and potential discrimination. Strategic directions for building trust are discussed, including transparency and explainability of AI decisions, independent testing and post-market monitoring, active involvement of medical professionals and patients in implementation, and

© Enikeev I.M., Solntsev V.I., Enikeev M.A., 2025

quality management based on international standards. The authors conclude that developing ‘calibrated’ trust, grounded in empirically proven benefits and fairness of AI application in clinical practice, is essential. The article also outlines priorities for further research tailored to the Russian context.

Key words: artificial intelligence, trust, healthcare, transparency, explainability, medical technology, regulation.

For citation: *Enikeev I.M., Solntsev V.I., Enikeev M.A. Trust in Artificial Intelligence Systems in Healthcare: Barriers and Strategies for Improvement // Lomonosov Public Administration Journal. Series 21. 2025. Vol. 22. № 4. P. 83–101.*

Received: 15.09.2025

Введение

Искусственный интеллект (ИИ) стремительно интегрируется в здравоохранение, открывая новые возможности — от повышения точности диагностики и персонализации терапии до оптимизации организационных процессов. На начало третьего квартала 2025 г. в США зарегистрировано более 1000 медицинских изделий с использованием ИИ, одобренных FDA для клинического применения¹. В Российской Федерации к середине 2024 года официально зарегистрировано 47 медицинских изделий с ИИ (из них 42 отечественных и 5 зарубежных), что подтверждено данными Росздравнадзора².

Несмотря на технологический прогресс и рост числа одобренных решений, в профессиональной среде и среди пациентов сохраняется настороженность по отношению к ИИ. Причина этого кроется не только в технических ограничениях или недостатке клинических данных, но прежде всего — в дефиците доверия, как к самим алгоритмам, так и к их разработчикам и организациям, внедряющим данные технологии³.

¹ U.S. Food and Drug Administration. Artificial Intelligence (AI) Enabled Medical Devices // FDA: офиц. Сайт [Электронный ресурс]. URL: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices> (дата обращения 18.07.2025).

² Зарегистрированные медицинские изделия с применением технологий искусственного интеллекта // ЕГИСЗ Минздрава России. [Электронный ресурс]. URL: <https://portal.egisz.rosminzdrav.ru/files/Зарегистрированные%20медицинские%20изделия%20с%20применением%20технологий%20искусственного%20интеллекта.pdf> (дата обращения 18.07.2025).

³ *Sagona M., Dai T., Macis M. et al. Trust in AI-assisted health systems and AI's trust in humans // npj Health Syst. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5* (дата обращения 18.07.2025).

Как справедливо отмечает экспертное сообщество, практически каждая дискуссия об ИИ в медицине в итоге сводится к вопросам доверия⁴. Доверие — фундаментальное понятие для медицины: пациент, начиная лечение, доверяет врачу, а врач — медицинским технологиям и данным. По определению социолога Н. Лумана, доверие позволяет принимать на себя риск в условиях неопределенности, открывая пространство для сотрудничества и принятия решений в сложных ситуациях⁵.

Однако специфика ИИ состоит в том, что многие современные алгоритмы функционируют как «черный ящик», не раскрывая пользователю внутреннюю логику работы, что усиливает скептицизм и подчеркивает необходимость обсуждения: достаточно ли прозрачности или важнее эмпирическая доказанность и безопасность? Сомнения усиливают риски ошибок, опасения предвзятости и опасность «отчуждения» решения от клинициста. Эти вопросы осложняются и тем, что пациенты не взаимодействуют с ИИ напрямую, а должны доверять врачам, что те компетентно используют технологию и не полагаются на нее вслепую.

Современная ситуация во многом напоминает промышленную революцию в Великобритании, когда опасения по поводу внедрения новых машин и автоматизации (движение луддитов) сменялись постепенным формированием новых практик доверия — как со стороны работников, так и общества⁶.

Таким образом, доверие к медицинским ИИ-системам — это сложная сеть взаимоотношений между пациентом, врачом, технологией и организацией⁷, в которой каждому из участников необходимо обоснованно полагаться на других.

В статье также рассмотрим, какие аспекты мешают установить прочное доверие к медицинскому ИИ, к чему приводит недостаток или избыток доверия, и какие шаги предпринимать для укрепления доверительных отношений в системе «пациент–врач–ИИ».

Материалы и методы

В работе проведен аналитический обзор современной научной литературы, международных стандартов и нормативных докумен-

⁴ Там же.

⁵ *Luhmann N.* Trust and Power. New York: John Wiley & Sons Inc., 1982.

⁶ The Editors of Encyclopaedia Britannica. Luddite // Encyclopaedia Britannica. URL: <https://www.britannica.com/event/Luddite> (дара обращения 18.07.2025).

⁷ *Sagona M., Dai T., Macis M. et al.* Trust in AI-assisted health systems and AI's trust in humans // *npj Health Syst.* 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

тов, посвященных вопросам доверия к системам ИИ в здравоохранении. Для анализа использовались публикации в ведущих зарубежных и российских рецензируемых журналах (NEJM AI, JAMA Network Open, International Journal of Medical Informatics и др.), официальные отчеты регуляторных органов (FDA, Росздравнадзор), а также результаты эмпирических исследований, отражающих мнения пациентов и медицинских специалистов. Особое внимание уделялось источникам, раскрывающим барьеры и успешные практики укрепления доверия к ИИ, включая опыт внедрения международных стандартов управления.

Восприятие пациентов: предпочтения и опасения

Результаты современных исследований свидетельствуют о неоднозначном отношении пациентов к внедрению искусственного интеллекта в медицинскую практику. Так, в работе J.W. Ayers и соавт. (2023) эксперты в слепом сравнении оценивали ответы на вопросы пациентов, подготовленные врачами и чат-ботом на основе ИИ (ChatGPT). Согласно полученным данным, 78,5% ответов ИИ были признаны «хорошими или очень хорошими», тогда как аналогичный показатель для врачей составил лишь 22,1%. Кроме того, ответы ИИ значительно превосходили врачебные по критерию эмпатии (45,1% против 4,6%)⁸. Таким образом, в экспериментальных условиях ответы ИИ воспринимались как более информативные и человечные, что подчеркивает актуальность дальнейших исследований в данной области.

В исследовании Duke University Health System также отмечена высокая удовлетворенность пациентов ответами, подготовленными с использованием ИИ: такие сообщения чаще воспринимались как содержательные и подробные. При этом было установлено, что явное раскрытие факта применения ИИ приводит к небольшому снижению уровня удовлетворенности и доверия (различия по 5-балльной шкале — 0,1–0,3 балла), однако общий уровень позитивных оценок сохраняется высоким (более 75% респондентов)⁹. Это указывает на то, что прозрачность использования ИИ допу-

⁸ Ayers J.W., Poliak A., Dredze M., et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum // JAMA Internal Medicine. 2023. Vol. 183(6). P. 589–596. DOI: 10.1001/jamainternmed.2023.1838 (дата обращения 18.07.2025).

⁹ Cavalier J.S., Goldstein B.A., Ravitsky V., et al. Ethics in Patient Preferences for Artificial Intelligence — Drafted Responses to Electronic Messages // JAMA Network Open. 2025. Vol. 8(3). Article e250449. DOI: 10.1001/jamanetworkopen.2025.0449 (дата обращения 18.07.2025).

стима, но требует взвешенного подхода для поддержания доверия пациентов.

Следует подчеркнуть, что отношение пациентов к применению искусственного интеллекта во многом определяется общим доверием к системе здравоохранения и медицинским организациям. Отдельного внимания заслуживает проблема распознавания пациентами источника медицинских рекомендаций. В работе Shekar и соавт. (2025) показано, что респонденты нередко не отличали ответы, подготовленные врачом, от сгенерированных большой языковой моделью. Более того, при сопоставимой точности респонденты оценивали ИИ-ответы не ниже, а в ряде случаев даже выше, чем ответы специалистов. Авторы отмечают, что участники эксперимента были склонны доверять даже тем рекомендациям ИИ, которые содержали неточности, что потенциально увеличивает риск следования ошибочным советам и нежелательных медицинских вмешательств¹⁰.

Недавний опрос взрослых жителей США выявил сравнительно низкий уровень доверия пациентов к медицинским организациям в вопросах ответственного применения ИИ и защиты от связанных с ним рисков: 66% респондентов не уверены в ответственности использования ИИ, а 58% сомневаются в своей защищенности от потенциального вреда. На доверие в первую очередь влияет общий опыт взаимодействия с системой здравоохранения: пациенты с исходно высоким уровнем доверия существенно чаще уверены в способности организации обеспечить безопасность при внедрении ИИ (отношение шансов 4,29; 95% ДИ: 3,25–5,67). В то же время опыт дискриминации в системе здравоохранения резко снижает уверенность в ответственном использовании ИИ (отношение шансов 0,66 и 0,57). При этом уровень знаний об ИИ и базовая цифровая грамотность статистически значимой роли не играют. Эти результаты подчеркивают, что ключевым фактором формирования доверия является не столько информирование о технологиях, сколько качество и прозрачность работы самой медицинской организации¹¹.

Таким образом, анализ современных данных свидетельствует о наличии у части пациентов как выраженного интереса и доверия к ИИ, так и сохраняющейся настороженности по поводу сниже-

¹⁰ Shekar S., Pataranutaporn P., Sarabu C., Cecchi G.A., Maes P. People Overtrust AI-Generated Medical Advice despite Low Accuracy // NEJM AI. 2025. Vol. 2(6) (дата обращения 18.07.2025).

¹¹ Longhurst C., Singh K., Chopra A., Atreja A., Brownstein J. A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness // NEJM AI. 2024. Vol. 1. DOI: 10.1056/AIpr2400223 (дата обращения 18.07.2025).

ния роли профессионального медицинского участия и возможных ошибок алгоритмов. Вероятно, оптимальной стратегией является применение ИИ под контролем медицинских специалистов, с обеспечением прозрачности и приоритетом безопасности пациентов.

Доверие медицинских специалистов

Одним из определяющих факторов успешного внедрения ИИ в здравоохранение является уровень доверия со стороны медицинских работников. Несмотря на признание потенциала ИИ для повышения эффективности диагностики и терапии, отношение профессионального сообщества остается неоднородным: энтузиазм инноваторов сочетается с настороженностью большинства специалистов, прежде всего из-за ограниченной доказательной базы и неочевидности реальных клинических выгод^{12,13}. Согласно международному исследованию Elsevier, более 95% врачей и исследователей осведомлены о современных ИИ-системах, однако лишь половина когда-либо применяла их в работе, а активное использование характерно только для 11% респондентов; при этом 72% специалистов ожидают существенного влияния ИИ на свою деятельность в ближайшие годы¹⁴.

Фундаментом доверия врача является убежденность в надежности и доказанности инструментов, их соответствии принципу «не навреди» и сохранении профессионального контроля за принятием решений^{15,16,17}. Как показано в недавних аналитических

¹² Insights 2024: Attitudes toward AI. Elsevier, July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

¹³ Longhurst C., Singh K., Chopra A., Atreja A., Brownstein J. A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness // NEJM AI. 2024. Vol. 1. DOI: 10.1056/AIp2400223 (дата обращения 18.07.2025).

¹⁴ Insights 2024: Attitudes toward AI. Elsevier, July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

¹⁵ Sagona M., Dai T., Macis M. et al. Trust in AI-assisted health systems and AI's trust in humans // npj Health Syst. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

¹⁶ Longhurst C., Singh K., Chopra A., Atreja A., Brownstein J. A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness // NEJM AI. 2024. Vol. 1. DOI: 10.1056/AIp2400223 (дата обращения 18.07.2025).

¹⁷ Goldberg C.B., Adams L., Blumenthal D., et al. RAISE Consortium. To Do No Harm — and the Most Good — with AI in Health Care // NEJM AI. 2024. 1(3). DOI: 10.1056/AIp2400036 (дата обращения 18.07.2025).

обзорах, несмотря на бурный рост числа публикаций о возможностях ИИ в медицине, лишь единичные модели прошли всестороннюю клиническую валидацию, и еще меньшее число — показали положительный эффект для реальных пациентов^{18,19}. Подобный разрыв между теорией и практикой закономерно порождает скептицизм: до тех пор, пока не будут получены убедительные данные о снижении числа ошибок или улучшении исходов лечения, доля сомневающихся среди медиков останется высокой. Критически важным направлением, как подчеркивают эксперты RAISE Consortium, становится обеспечение безопасности и пользы для пациента как основного критерия внедрения ИИ в клиническую практику²⁰.

Значимым фактором доверия выступает и безопасность: медики опасаются сбоев ИИ, которые могут привести к ложным тревогам или пропуску опасных состояний^{21,22}. В профессиональном сообществе регулярно обсуждаются случаи неудачного внедрения ИИ, что формирует дополнительные барьеры. Немаловажно и то, каким образом система интегрируется в рабочий процесс: если внедрение ИИ проходит без согласования с коллективом и должного обучения, это усиливает сопротивление персонала²³. Врачи ценят сохранение автономии и контроль над процессом принятия решений; доверие к алгоритму возрастает, когда он прозрачен, адаптируется к клиническим задачам и дополняет, а не подменяет экспертизу врача.

¹⁸ Longhurst C., Singh K., Chopra A., Atreja A., Brownstein J. A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness // *NEJM AI*. 2024. Vol. 1. DOI: 10.1056/AIp2400223 (дата обращения 18.07.2025).

¹⁹ Huang J., Wittbrodt M.T., Teague C.N., et al. Efficiency and Quality of Generative AI-Assisted Radiograph Reporting // *JAMA Network Open*. 2025. Vol. 8(6). Article e2513921. DOI: 10.1001/jamanetworkopen.2025.13921 (дата обращения 18.07.2025).

²⁰ Goldberg C.B., Adams L., Blumenthal D., et al. RAISE Consortium. To Do No Harm — and the Most Good — with AI in Health Care // *NEJM AI*. 2024. 1(3). DOI: 10.1056/AIp2400036 (дата обращения 18.07.2025).

²¹ Sagona M., Dai T., Macis M. et al. Trust in AI-assisted health systems and AI's trust in humans // *npj Health Syst*. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

²² Goldberg C.B., Adams L., Blumenthal D., et al. RAISE Consortium. To Do No Harm — and the Most Good — with AI in Health Care // *NEJM AI*. 2024. 1(3). DOI: 10.1056/AIp2400036 (дата обращения 18.07.2025).

²³ Insights 2024: Attitudes toward AI. Elsevier. July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BW RibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

Результаты недавних рандомизированных исследований свидетельствуют, что, несмотря на высокую оценку эффективности ИИ-систем для генерации медицинских заключений (например, при рентгенологических исследованиях), специалисты все равно склонны доверять итоговым выводам только после личной проверки и готовы учитывать алгоритмические ошибки²⁴. При этом пациенты, особенно в ситуациях высокого риска, могут выражать меньше доверия к врачу, если тот использует решения, поддержанные ИИ, что особенно актуально для женщин и групп с низким исходным доверием к технологиям²⁵.

Особое место занимает вопрос компетентности: молодые специалисты, имеющие опыт работы с цифровыми технологиями, как правило, быстрее адаптируются к новым инструментам и проявляют большую готовность к их использованию²⁶. Вместе с тем наставничество и двусторонний обмен опытом между поколениями способствуют формированию сбалансированного отношения к ИИ и снижению риска как чрезмерного, так и недостаточного доверия.

Обобщая, можно заключить, что ключевыми условиями роста доверия специалистов к медицинским ИИ являются накопление эмпирически подтвержденных данных, системное обучение, постоянный этический надзор и сохранение приоритета клинической экспертизы в принятии решений.

Общие риски и угрозы внедрения ИИ

Реализация ИИ в здравоохранении сопряжена с целым рядом системных рисков, существенно влияющих на уровень доверия медицинских работников и пациентов. К ключевым угрозам, выявленным в современных исследованиях, относятся: утечка или несанкционированная обработка персональных данных, отсутствие прозрачности («черный ящик»), неясность распределения ответственности между врачом и системой, а также предвзятость

²⁴ Huang J., Wittbrodt M.T., Teague C.N., et al. Efficiency and Quality of Generative AI-Assisted Radiograph Reporting // JAMA Network Open. 2025. Vol. 8(6). Article e2513921. DOI: 10.1001/jamanetworkopen.2025.13921 (дата обращения 18.07.2025).

²⁵ Zondag A.G.M., Rozestraten R., Grimmelikhuijsen S.G., et al. The Effect of Artificial Intelligence on Patient-Physician Trust: Cross-Sectional Vignette Study // J Med Internet Res. 2024. Vol. 26. Article e50853. DOI: 10.2196/50853 (дата обращения 18.07.2025).

²⁶ Insights 2024: Attitudes toward AI. Elsevier. July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

алгоритмов, ведущая к диагностическим ошибкам и дискриминации. Сохраняются и проблемы гипо- и гипердиагностики, уязвимости кибербезопасности, технических сбоев, а также юридические пробелы в регулировании новых технологий. Как видно из обобщающей Таблицы 1, каждый из этих факторов способен привести к снижению безопасности, качества помощи и доверия участников процесса к системам искусственного интеллекта²⁷.

Таблица 1

Основные категории рисков и угроз при внедрении ИИ в здравоохранении²⁸

Категория риска	Описание	Пример проявления
Конфиденциальность	Утечка, несанкционированная обработка персональных данных	Взлом базы, передача изображений на внешний сервер
Этическая ответственность	Неясность ответственности за решение ИИ, снижение роли врача	Алгоритм не выявил болезнь, неясно, кто отвечает; потеря навыков у врачей
Недоверие пациентов	Скептицизм к решениям, принятым алгоритмом	Пациент отказывается от лечения, снижение удовлетворенности
«Черный ящик»	Непрозрачность, невозможность объяснить логику решений ИИ	Врач не понимает, на чем основано решение
Предвзятость алгоритма	Смещение обучающих данных, ошибки или дискриминация	Модель хуже диагностирует у представителей разных рас, пропуск редких патологий
Гипо-/гипердиагностика	Ложноотрицательные/положительные результаты	Пропуск опухоли (ложно-негатив), избыток подозрений (ложно-позитив)
Сбои, кибербезопасность	Технические отказы, атаки, ошибки программного обеспечения	Технические сбои в работе СШКР, искажение данных злоумышленниками
Юридические пробелы	Недостаточность нормативной базы для применения ИИ	Нет протокола по оформлению решений ИИ, сложность защиты в суде

²⁷ Khan M.M., Shah N., Shaikh N., et al. Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges // International Journal of Medical Informatics. 2025. Vol.195. Article 105780. DOI: 10.1016/j.ijmedinf.2024.105780 (дата обращения 18.07.2025).

²⁸ Разработано автором.

Риски избыточного доверия

Помимо угроз, связанных с недостатком доверия к ИИ, все большую актуальность приобретают риски чрезмерной уверенности в алгоритмах. Согласно глобальным опросам, большинство специалистов связывают с ИИ ожидания ускорения поиска знаний, расширения исследований и снижения издержек, однако при этом 94% респондентов опасаются дезинформации, а 86% — критических ошибок²⁹. В клинической практике избыточное доверие проявляется феноменом автоматизации ошибок: когда медицинский работник или пациент без должной критичности следует рекомендациям системы поддержки решений (automation bias), что может привести к неправильным вмешательствам и неблагоприятным исходам³⁰. В качестве реального примера, A. Zaoui и соавт. (2025) описывают случай, когда пациент, получив рекомендацию от ChatGPT, самостоятельно попытался провести медицинскую манипуляцию (лигирование геморроидальных узлов), что привело к осложнениям и необходимости экстренной медицинской помощи. Авторы подчеркивают, что отсутствие критического мышления и переоценка возможностей ИИ могут не только ввести в заблуждение, но и нанести реальный вред здоровью, а потому ответы ИИ всегда должны подтверждаться профессионалом³¹. В организационном аспекте чрезмерное полагание на алгоритмы без резервных проверок повышает уязвимость системы при сбоях и новых сценариях, что требует обязательного присутствия человеческого контроля и корректирующих механизмов (см. Таблицу 1). Формирование доверия должно строиться на эмпирически подтвержденных данных и поэтапном внедрении, что позволит минимизировать угрозы и сохранить безопасность пациентов³².

²⁹ Insights 2024: Attitudes toward AI. Elsevier, July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

³⁰ Sagona M., Dai T., Macis M. et al. Trust in AI-assisted health systems and AI's trust in humans // npj Health Syst. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

³¹ Zaoui A., Aloulou H., Nouri M., et al. Misuse of Artificial Intelligence in Medical Practice: A Case Report // Clinical Practice & Cases in Emergency Medicine. 2025. Vol. 9(3). P. 196–199. DOI: 10.5811/cpcem.2024.11.64317 (дата обращения 18.07.2025).

³² Sagona M., Dai T., Macis M. et al. Trust in AI-assisted health systems and AI's trust in humans // npj Health Syst. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

Пути укрепления доверия к медицинскому ИИ

Преодоление существующих барьеров и формирование обоснованного доверия к системам искусственного интеллекта (ИИ) в здравоохранении требует комплексного подхода, сочетающего технологические, управленческие и этические меры. Опыт международных исследований и рекомендации ведущих экспертов в области цифровой трансформации медицины свидетельствуют, что эффективность подобных стратегий во многом определяется системностью внедрения, прозрачностью процессов и участием всех заинтересованных сторон^{33,34,35}.

Системная диагностика и аудит процессов

Перед внедрением ИИ целесообразно проводить детальный анализ бизнес-процессов и инфраструктуры медицинской организации. Международный стандарт ISO 20700:2017 «Guidelines for management consultancy services»³⁶ рекомендует выделять следующие этапы^{37,38}:

- Диагностика и анализ исходных процессов (ассесмент текущей цифровой зрелости, выявление ключевых точек риска и уязвимости в отношении доверия);
- Постановка задач с учетом конкретных клинических и управленческих целей;
- Аналитическая валидация исходных данных, включая аудит источников и структуру их сбора.

³³ Insights 2024: Attitudes toward AI. Elsevier, July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

³⁴ *Sagona M., Dai T., Macis M. et al.* Trust in AI-assisted health systems and AI's trust in humans // *npj Health Syst.* 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

³⁵ *Rosenbacke R., Melhus Å., McKee M., Stuckler D.* How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Healthcare: a Systematic Review // *JMIR Preprints.* 2023. URL: <https://preprints.jmir.org/preprint/53207> (дата обращения 18.07.2025).

³⁶ International Organization for Standardization. ISO 20700:2017. Guidelines for management consultancy services. Geneva: ISO. 2017. URL: <https://www.iso.org/standard/63501.html> (дата обращения 18.07.2025).

³⁷ *Sagona M., Dai T., Macis M. et al.* Trust in AI-assisted health systems and AI's trust in humans // *npj Health Syst.* 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

³⁸ Insights 2024: Attitudes toward AI. Elsevier, July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

В ряде случаев возможно применение модели оценки траекторий внедрения — выбора оптимальных комбинаций алгоритмов и организационных решений в зависимости от профиля рисков, структуры пациента и организационной среды^{39,40}.

Эмпирическая проверка эффективности и безопасности

Одним из главных факторов доверия является наличие убедительной доказательной базы по реальным клиническим и организационным результатам применения ИИ. На практике это означает^{41,42}:

- Проведение проспективных и ретроспективных исследований с участием реальных пациентов, сравнение с действующими стандартами;
- Организация многоцентровых пилотных проектов с прозрачным протоколом и регистрацией всех исходов;
- Оценка показателей надежности, воспроизводимости и справедливости алгоритмов на разных подгруппах пациентов.

Стандартизированные испытания и независимый аудит критически важны для принятия решений о дальнейшем масштабировании технологии.

Прозрачность, объяснимость и коммуникация

Техническая и организационная прозрачность — важнейший фактор укрепления доверия как среди врачей, так и среди пациентов. Современные методы объяснимого ИИ (XAI, explainable AI) позволяют делать ключевые решения алгоритма понятными и сопоставимыми с клинической логикой. Однако исследования показывают, что избыточно сложные или формальные объяснения

³⁹ *Sagona M., Dai T., Macis M. et al.* Trust in AI-assisted health systems and AI's trust in humans // *npj Health Syst.* 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

⁴⁰ *Coiera E., Fraile-Navarro D.* AI as an Ecosystem — Ensuring Generative AI Is Safe and Effective // *NEJM AI.* 2024. Vol. 1(9). DOI: 10.1056/AIip2400611 (дата обращения 18.07.2025).

⁴¹ *Sagona M., Dai T., Macis M. et al.* Trust in AI-assisted health systems and AI's trust in humans // *npj Health Syst.* 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

⁴² *Insights 2024: Attitudes toward AI.* Elsevier, July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibynJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

могут не повысить, а даже снизить доверие, особенно если они не подкреплены эмпирическими данными⁴³.

Для укрепления доверия необходимы^{44,45}:

- Публичные протоколы валидации и объяснения алгоритмов для специалистов;
- Доступное информирование пациентов о применении ИИ, возможностях и ограничениях системы;
- Четкая коммуникация о рисках, механизмах контроля и правах пациента на пересмотр решения.

Справедливость, этичность и учет разнообразия

ИИ должен демонстрировать одинаковую надежность и безопасность для всех подгрупп населения. Это достигается за счет^{46,47}:

- Обязательного тестирования и валидации на репрезентативных выборках с учетом этнических, социальных и гендерных различий;
- Внедрения протоколов оценки и устранения алгоритмических bias;
- Создания независимых экспертных советов и комитетов по этике цифрового здравоохранения.

Учитывая историю недоверия со стороны меньшинств и социально уязвимых групп, прозрачность и этичность разработки и внедрения должны быть подкреплены регулярными аудитам и привлечением представителей сообщества⁴⁸.

⁴³ Rosenbacke R., Melhus Å., McKee M., Stuckler D. How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Healthcare: a Systematic Review // JMIR Preprints. 2023. URL: <https://preprints.jmir.org/preprint/53207> (дата обращения 18.07.2025).

⁴⁴ Insights 2024: Attitudes toward AI. Elsevier. July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BW RibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

⁴⁵ Rosenbacke R., Melhus Å., McKee M., Stuckler D. How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Healthcare: a Systematic Review // JMIR Preprints. 2023. URL: <https://preprints.jmir.org/preprint/53207> (дата обращения 18.07.2025).

⁴⁶ Sagona M., Dai T., Macis M. et al. Trust in AI-assisted health systems and AI's trust in humans // npj Health Syst. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

⁴⁷ Nong P., Platt J. Patients' Trust in Health Systems to Use Artificial Intelligence // JAMA Network Open. 2025. Vol. 8(2). Article e2460628. DOI: 10.1001/jama-networkopen.2024.60628 (дата обращения 18.07.2025).

⁴⁸ Sagona M., Dai T., Macis M. et al. Trust in AI-assisted health systems and AI's trust in humans // npj Health Syst. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

Обучение, вовлечение и поддержка кадров

Уровень доверия напрямую зависит от цифровой грамотности и вовлеченности медицинского персонала, а также их понимания логики и ограничений ИИ⁴⁹:

- Проведение регулярных образовательных программ, тренингов и семинаров;
- Вовлечение врачей, пациентов и ИТ-специалистов в процесс совместного проектирования и выбора ИИ-решений;
- Создание программ наставничества и обратной связи по использованию ИИ в реальной клинической практике.

Управление и сертификация процессов (ISO 20700)

Ключевой особенностью современных программ внедрения ИИ является институционализация процессов управления качеством и безопасностью^{50, 51}:

- Назначение ответственных за развитие и контроль использования ИИ (например, Chief Health AI Officer (заместитель руководителя по ИИ и цифровой медицине)⁵²);
- Внедрение корпоративных регламентов и политик по безопасности, этике, прозрачности и ответственности;
- Регулярная сертификация внутренних процессов и внешняя оценка по международным стандартам (например, ISO 20700);
- Обязательная переоценка и переподтверждение качества и безопасности решений на каждом этапе жизненного цикла алгоритма.

⁴⁹ Insights 2024: Attitudes toward AI. Elsevier. July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

⁵⁰ *Sagona M., Dai T., Macis M. et al.* Trust in AI-assisted health systems and AI's trust in humans // *npj Health Syst.* 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5 (дата обращения 18.07.2025).

⁵¹ Insights 2024: Attitudes toward AI. Elsevier. July 2024. 43 p. URL: https://assets.ctfassets.net/o78em1y1w4i4/6BWRibyJNQLYkKWwKw7SVf/64c04b53ca9cc0795ac811f583f7eebb/Insights_2024_Attitudes_To_AI_Full_Report.pdf (дата обращения 18.07.2025).

⁵² *Beecy A.N., Longhurst C.A., Singh K., Wachter R.M., Murray S.G.* The Chief Health AI Officer — An Emerging Role for an Emerging Technology // *NEJM AI.* 2024. Vol. 1(7). DOI: 10.1056/AIp2400109 (дата обращения 18.07.2025).

Контроль и мониторинг, обеспечение обратной связи

Для устойчивого доверия требуется:

- Постоянный мониторинг работы ИИ-систем, автоматизированный сбор и анализ ошибок, обращений, неблагоприятных исходов;
- Внедрение механизмов быстрого реагирования и корректировки алгоритмов;
- Публичные отчеты о работе систем и результатах внешнего аудита.

Чек-лист инструментов и мер

Для практического применения перечисленных принципов может быть использован краткий чек-лист инструментов и мер по укреплению доверия к ИИ, обобщенный на основании вышеперечисленных рекомендаций (см. табл. 2).

Таблица 2

Чек-лист инструментов и мер⁵³

Этап внедрения	Ключевые действия
1. Диагностика и аудит	Оценка процессов, анализ цифровой зрелости, выявление уязвимостей
2. Проверка эффективности	Клинические и организационные исследования, пилотные проекты
3. Прозрачность и объяснимость	Протоколы ХАИ, открытые разъяснения для врачей и пациентов
4. Справедливость и этика	Валидация на разных популяциях, мониторинг bias, этические комитеты
5. Обучение и вовлечение	Образовательные программы, совместный выбор и настройка решений
6. Управление и сертификация	Ответственные лица, внутренние регламенты, внешняя сертификация
7. Мониторинг и обратная связь	Непрерывный контроль, анализ ошибок, корректирующие меры

Будущие исследования

Большинство данных о доверии к медицинскому ИИ получено за рубежом и отражает особенности иностранных систем здравоохранения. В российских условиях такие исследования пока еди-

⁵³ Разработано авторами.

ничны, что ограничивает возможность учета локальных факторов. В дальнейшем необходимы национальные опросы и анализ барьеров, чтобы адаптировать эффективные меры по укреплению доверия к ИИ с учетом отечественной специфики.

Заключение

Доверие к медицинскому ИИ не возникает «по умолчанию» и не должно быть ни слепым, ни излишне скептическим. Наша цель — **калиброванное доверие**: ровно настолько высокое, чтобы использовать сильные стороны технологий, и достаточно осторожное, чтобы не игнорировать их ограничения.

Это доверие опирается на четыре столпа:

1. **Доказанная клиническая польза в реальной практике**, а не только высокие метрики на ретроспективных датасетах. Нужны проспективные исследования и постоянный пост-маркетинговый мониторинг.
2. **Прозрачность и подотчетность**: понятные правила использования, раскрытие ограничений, объяснимость там, где она помогает клиницисту и пациенту принимать решения.
3. **Справедливость и отсутствие дискриминации**: обязательные проверки качества алгоритмов на разных группах пациентов, регулярные fairness-аудиты.
4. **Управление и ответственность**: четкие роли, процессы, политики (включая независимый контроль качества, этические комитеты, периодическую переоценку моделей).

Следующие практические шаги очевидны: провести **репрезентативные национальные опросы** пациентов и врачей о доверии к ИИ; запустить **публичные реестры результатов работы алгоритмов** (точность, ошибки, корректировки); закрепить **регулярную переоценку и мониторинг** моделей на протяжении всего их жизненного цикла.

Только так ИИ станет инструментом, которому **обоснованно доверяют**, который усиливает, а не подменяет клиническую экспертизу, и реально улучшает исходы лечения, не подрывая права пациента и профессиональную автономию врача.

Литература

Ayers J.W., Poliak A., Dredze M., et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum // JAMA Internal Medicine. 2023. Vol. 183(6). P. 589–596. DOI: 10.1001/jamainternmed.2023.1838

Beecy A.N., Longhurst C.A., Singh K., Wachter R.M., Murray S.G. The Chief Health AI Officer — An Emerging Role for an Emerging Technology // *NEJM AI*. 2024. Vol. 1(7). DOI: 10.1056/AIp2400109

Cavalier J.S., Goldstein B.A., Ravitsky V., et al. Ethics in Patient Preferences for Artificial Intelligence—Drafted Responses to Electronic Messages // *JAMA Network Open*. 2025. Vol. 8(3). Article e250449. DOI: 10.1001/jamanetworkopen.2025.0449

Coiera E., Fraile Navarro D. AI as an Ecosystem — Ensuring Generative AI Is Safe and Effective // *NEJM AI*. 2024. Vol. 1(9). DOI: 10.1056/AIp2400611

Goldberg C.B., Adams L., Blumenthal D., et al. RAISE Consortium. To Do No Harm — and the Most Good — with AI in Health Care // *NEJM AI*. 2024. Vol. 1(3). DOI: 10.1056/AIp2400036

Huang J., Wittbrodt M.T., Teague C.N., et al. Efficiency and Quality of Generative AI — Assisted Radiograph Reporting // *JAMA Network Open*. 2025. Vol. 8(6): Article e2513921. DOI: 10.1001/jamanetworkopen.2025.13921

Khan M.M., Shah N., Shaikh N., et al. Towards secure and trusted AI in health-care: A systematic review of emerging innovations and ethical challenges // *International Journal of Medical Informatics*. 2025. Vol. 195. Article 105780. DOI: 10.1016/j.ijmedinf.2024.105780

Longhurst C., Singh K., Chopra A., Atreja A., Brownstein J. A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness // *NEJM AI*. 2024. Vol. 1. DOI: 10.1056/AIp2400223

Luhmann N. Trust and Power. New York: John Wiley & Sons Inc., 1982.

Nong P., Platt J. Patients' Trust in Health Systems to Use Artificial Intelligence // *JAMA Network Open*. 2025. Vol. 8(2). Article e2460628. DOI: 10.1001/jamanetworkopen.2024.60628

Rosenbacke R., Melhus Å., McKee M., Stuckler D. How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health-care: a Systematic Review // *JMIR Preprints*. 2023. URL: <https://preprints.jmir.org/preprint/53207>

Sagona M., Dai T., Macis M., et al. Trust in AI assisted health systems and AI's trust in humans // *npj Health Syst*. 2025. Vol. 2. Article 10. DOI: 10.1038/s44401-025-00016-5

Shekar S., Pataranutaporn P., Sarabu C., Cecchi G.A., Maes P. People Overtrust AI Generated Medical Advice despite Low Accuracy // *NEJM AI*. 2025. Vol. 2(6).

Zaoui A., Aloulou H., Nouri M., et al. Misuse of Artificial Intelligence in Medical Practice: A Case Report // *Clinical Practice & Cases in Emergency Medicine*. 2025. Vol. 9(3). P. 196–199. DOI: 10.5811/cpcem.2024.11.64317

Zondag A.G.M., Rozestraten R., Grimmelikhuijsen S.G., et al. The Effect of Artificial Intelligence on Patient Physician Trust: Cross Sectional Vignette Study // *J Med Internet Res*. 2024. Vol. 26. Article e50853. DOI: 10.2196/50853

ИНФОРМАЦИЯ ОБ АВТОРАХ:

Еникеев Игорь Маратович — заместитель директора Департамента развития цифрового продукта и искусственного интеллекта «ГЭОТАР-Цифра», ООО «Консультант студента», Москва, Россия; *e-mail*: enikeevigor99@gmail.com

Солнцев Виктор Игоревич — к.т.н., генеральный директор НП «ИВМ Консалтинг Групп», Москва, Россия; *e-mail*: vi.solntsev@gmail.com

Еникеев Марат Альбертович — к.м.н., главный врач ГБУЗ РК «Керченская больница № 1 имени Н.И. Пирогова», Керчь, Россия; *e-mail*: marat_misha@mail.ru

ABOUT THE AUTHORS:

Enikeev Igor Maratovich — Deputy Director of the Department of Digital Product Development and Artificial Intelligence, “GEOTAR-Digital”, LLC “Student Consultant”, Moscow, Russia; *e-mail*: enikeevigor99@gmail.com

Solntsev Viktor Igorevich — PhD (Tech.), Associate Professor, General Director of NP “IVM Consulting Group”, Certified Management Consultant (CMC), Moscow, Russia; *e-mail*: vi.solntsev@gmail.com

Enikeev Marat Albertovich — PhD (Med.), Chief Physician, State Budgetary Healthcare Institution of the Republic of Crimea “Kerch Hospital No. 1 named after N.I. Pirogov”, Kerch, Russia; *e-mail*: marat_misha@mail.ru