DOI: https://doi.org/10.17816/KMJ690475 EDN: LMKXIK

Переосмысление биоэтических принципов в эпоху искусственного интеллекта: вызовы для автономии, справедливости и благодеяния в медицине

Ф.Т. Нежметдинова[1], М.Э. Гурылева[2]

[1] Казанский государственный аграрный университет, г. Казань, Россия;
[2] Казанский государственный медицинский университет, г. Казань, Россия

АННОТАЦИЯ

Широкое внедрение алгоритмов искусственного интеллекта (ИИ) в клиническую практику — от диагностики заболеваний до роботизированной хирургии — ставит под сомнение адекватность традиционных биоэтических принципов, сформулированных в контексте принятия решений человеком-врачом. Настоящая статья направлена на критический анализ применимости принципов Бичампа — уважения автономии, благодеяния и справедливости — к реалиям медицины, опосредованной ИИ, и предлагает конкретные пути их адаптации. Методология исследования включает системный обзор и тематический анализ международной научной литературы, клинических случаев и нормативных документов за период 2015–2023 гг. В результате выявлены фундаментальные противоречия: принцип благодеяния сталкивается с проблемой «чёрного ящика» и размывания ответственности; принцип автономии требует пересмотра моделей информированного согласия и закрепления «права на объяснение»; принцип справедливости нарушается из-за алгоритмической предвзятости; конфиденциальность данных требует новых подходов, таких как федеративное обучение. В заключение подчёркивается, что сохранение доверия к медицине требует не отказа от традиционных принципов биоэтики, а их эволюции — через дополнение прозрачностью, подотчётностью и технической справедливостью. Это предполагает разработку новых регуляторных стандартов, обязательный аудит алгоритмов и интеграцию этического дизайна в процесс создания медицинских систем ИИ.

Ключевые слова: биоэтика; искусственный интеллект; принципы Бичампа; информированное согласие; цифровое здравоохранение.

Как цитировать:
Нежметдинова Ф.Т., Гурылева М.Э. Переосмысление биоэтических принципов в эпоху искусственного интеллекта: вызовы для автономии, справедливости и благодеяния в медицине // Казанский медицинский журнал. 2025. DOI: 10.17816/KMJ690475 EDN: LMKXIK

Рукопись получена: 16.09.2025 **Рукопись одобрена:** 07.10.2025 **Опубликована online:** 11.11.2025

DOI: https://doi.org/10.17816/KMJ690475 EDN: LMKXIK

Reconsidering Bioethical Principles in the Era of Artificial Intelligence: Challenges for Autonomy, Justice, and Beneficence in Medicine

Farida T. Nezhmetdinova[1], Marina E. Guryleva[2]

[1] Kazan State Agrarian University, Kazan, Russia;
[2] Kazan State Medical University, Kazan, Russia

ABSTRACT

The widespread integration of artificial intelligence (AI) algorithms into clinical practice, from disease diagnosis to robotic surgery, raises questions about the adequacy of traditional bioethical principles developed for human physician decision-making. This article aimed to assess the applicability of the Beauchamp principles, namely, respect for autonomy, beneficence, and justice, to the realities of AI-mediated medicine and to propose specific strategies for their adaptation. The study involved a systematic review and thematic analysis of international scientific data, clinical cases, and regulatory documents published between 2015 and 2023. The analysis revealed fundamental contradictions. The principle of beneficence is challenged by the black box problem and diffusion of responsibility; autonomy requires revision of informed consent models and codification of the right to explanation; justice is undermined by algorithmic bias; and data confidentiality demands new approaches such as federated learning. Therefore, maintaining trust in medicine requires not the rejection but the evolution of traditional bioethical principles through the incorporation of transparency, accountability, and technical fairness. This indicates the need for new regulatory standards, mandatory algorithmic audits, and the integration of ethical design into the creation of AI-based medical systems.

Keywords: bioethics; artificial intelligence; Beauchamp principles; informed consent; digital health.

To cite this article:
Nezhmetdinova FT, Guryleva ME. Reconsidering bioethical principles in the era of artificial intelligence: challenges for autonomy, justice, and beneficence in medicine. *Kazan Medical Journal.* 2025. DOI: 10.17816/KMJ690475 EDN: LMKXIK

Submitted: 16.09.2025 **Accepted:** 07.10.2025 **Published online:** 11.11.2025

ВВЕДЕНИЕ

Глобальный рынок медицинских технологий, основанный на искусственном интеллекте (ИИ), обеспечивает устойчивый и экспоненциальный рост. Современные алгоритмы машинного и глубокого обучения уже превосходят медицинские знания в расширении ряда диагностических задач: в частности, анализ гистологических препаратов в онкологической практике; эквивалентные изображения, методы медицинской визуализации (например, возникновение диабетической ретинопатии по фундус-фотографиям), а также в реализации профилактических заболеваний (включая прогнозирование признаков острого повреждения за 48 ч до клинических проявлений). Тем не менее широкомасштабное внедрение ИИ-системы в клинической практике связано не только с техническими, но и с фундаментальными этико-правовыми вызовами.

Традиционная биоэтика, сформированная в рамках гуманистической парадигмы, предполагает, что субъектом моральной ответственности и этического суждения является человек — врач, исследователь или иной медицинский работник, обладающий возможностями к эмпатии, рефлексии и вербализации обоснования своих решений. В условиях применения ИИ, особенно архитектуры глубокого обучения, возникает принципиальная особенность: такие системы функционируют как «чёрные ящики», где между входными данными и итоговым выводом отсутствует интерпретируемый, прослеживаемый и объяснимый логический путь. Такое обстоятельство создаёт существенный разрыв между динамикой технологического развития и устоявшимися этико-правовыми нормами, что, в свою очередь, создаёт риски дальнейшего углубления данного диссонанса в будущем.

Внедрение ИИ в медицину — не просто технологическое обновление, фундаментальный сдвиг в парадигме клинического мышления, принятия решений и распределения ответственности. В среднесрочной перспективе (5–15 лет), когда ИИ-системы перейдут от пилотных проектов к рутинной работе в диагностических, прогностических и терапевтических процессах, их влияние на биоэтические нормы станет системным и необратимым.

Цель обзора — ретроспективный взгляд на эволюцию содержания, стандартов и практических методов биоэтики под влиянием ИИ и адаптацию традиционных биоэтических подходов в условиях методической медицины.

Задачи исследования:

1. На основе конкретных примеров показать, как применение ИИ сегодня соотносится с международными стандартами базовых биоэтических механизмов (благодеяние, автономия пациента, справедливость).

2. Выявить и обсудить правовые и технические коллизии, возникающие в результате ошибок, допущенных ИИ-системами, включая вопросы ответственности, прозрачности и подотчётности.

Разработать и обосновать предложения по адаптации и модификации существующих биоэтических принципов, методических указаний и регуляторных подходов с учётом специфики современной «алгоритмической медицины».

Таким образом, исследование направлено на преодоление концептуального разрыва между быстрым технологическим прогрессом и медленной эволюцией этико-правовых норм, что приводит к теоретико-методологической основе для ответственного и этически обоснованного развития ИИ в здравоохранении.

Настоящее исследование представляет собой комплексный структурированный обзор и концептуальный анализ. Использованы методы тематического синтеза и критического анализа, проведён сравнительный анализ этических и правовых международных и отечественных регламентов, а также применён кейс-метод (метод анализа ситуаций). Стандарты биоэтики традиционно закрепляются в виде деклараций (Хельсинкская декларация, Лиссабонская декларация, Белмонтский отчёт) [1], профессиональных кодексов и институциональных протоколов.

РЕГУЛИРОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РОССИЙСКОЙ ФЕДЕРАЦИИ

В Российской Федерации существует правовая и методологическая база регламентации применения ИИ[1,2] [2–4]. Однако, как отмечают многие исследователи, ИИ требует гибких, адаптированных и контекстно-зависимых стандартов[3,4] [4–6], поскольку:

• алгоритмы ИИ постоянно изменяются в процессе обучения, что делает невозможным статическое регулирование данной сферы;

[1] Национальная стратегия развития искусственного интеллекта на период до 2030 года. Утверждена Указом Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации». Режим доступа: http://kremlin.ru/acts/bank/44731 Дата обращения: 01.05.2025.

[2] Приказ Федерального агентства по техническому регулированию и метрологии (Росстандарт) от 25.07.2019 № 1732 (ред. от 20.01.2021) «О создании технического комитета по стандартизации "Искусственный интеллект"». Режим доступа: https://www.rst.gov.ru/portal/gost/home/activity/documents/orders#/order/104460 Дата обращения: 01.05.2025.

[3] Альянс в сфере искусственного интеллекта. Кодекс этики в сфере ИИ [интернет]. © AI Alliance Russia. Режим доступа: https://ethics.a-ai.ru/ Дата обращения: 22.02.2025.

[4] Центр искусственного интеллекта НИУ ВШЭ. Этика искусственного интеллекта [интернет]. Режим доступа: https://cs.hse.ru/aicenter/ethics Дата обращения: 22.02.2025.

• клинический контекст применения ИИ варьирует от рутинной диагностики до экстренной помощи, требующей разных уровней контроля;

• этические нормы применения ИИ должны контролироваться на всех этапах: не только при получении результата, но и в процессе принятия решений (включая прозрачность действий, понятность алгоритмов принятия решений и возможность оспаривания выводов, сформулированных с помощью ИИ).

ЭТИЧЕСКИЕ ПРИНЦИПЫ

Ключевые этические принципы, составляющие фундамент современной биоэтической парадигмы в медицине, — принцип благодеяния, принцип уважения автономии личности, принцип справедливости и принцип конфиденциальности — представляют собой систематизированную нормативную структуру, известную как принципиализм.

Теоретическое оформление и консолидация данной системы в её современном виде принадлежат американским философам Т.Л. Бичампу (T.L. Beauchamp) и Дж.Ф. Чайлдрессу (J.F. Childress). Их фундаментальный совместный труд «Принципы биомедицинской этики» (Principles of Biomedical Ethics) [7], впервые опубликованный в 1979 году, выступил в качестве катализатора для унификации языка и методологии анализа моральных дилемм в здравоохранении. Изначальная концепция авторов оперировала четырьмя базовыми категориями: автономия, непричинение вреда, благодеяние и справедливость. Принцип конфиденциальности, хотя и не всегда выделяемый как первичный, логически выводится из фундаментального требования уважения автономии пациента, поскольку защита приватности персональной информации является необходимым условием для реализации самоопределения личности.

Возникновение принципиализма не было инновационным в строгом смысле, но являлось результатом критического синтеза и адаптации ключевых направлений западной философско-этической мысли.

Философская основа принципа благодеяния и, отчасти, справедливости укоренена в традиции утилитаризма (Джереми Бентам, Джон Стюарт Милль), где моральная ценность действия определяется его последствиями, а именно достижением максимального блага для максимального числа людей [8].

В противовес этому, кантианская деонтология (Иммануил Кант) заложила фундамент для принципа уважения автономии. Кантовский категорический императив, требующий рассматривать человека всегда как цель и никогда лишь как средство, обеспечил философское обоснование для современных концепций информированного согласия и признания самоценности личности пациента. Абсолютизация запрета на причинение ущерба, имплицитно присутствующая в деонтологии, также напрямую питает принцип непричинения вреда [8].

Теория справедливости, сформулированная Джоном Ролзом в его фундаментальной работе «Теория справедливости» (A Theory of Justice, 1971), существенно повлияла на концептуальное содержание биоэтического принципа справедливости. Сформулированный Ролзом «принцип различия», допускающий социальное и экономическое неравенство лишь в том случае, если оно приносит наибольшую пользу наименее привилегированным членам общества, предоставил мощный нормативный инструментарий для анализа проблем справедливого распределения дефицитных медицинских ресурсов и обеспечения равного доступа к медицинской помощи [9].

Непосредственным практическим предшественником работы Бичампа и Чайлдресса выступил «Отчёт Белмонта» (The Belmont Report, 1979) — основополагающий документ, разработанный для защиты прав и благополучия участников научных исследований [10]. В Отчёте были определены три базовых этических принципа: уважение к личности, благодеяние и справедливость. Данная модель, таким образом, развила и дополнила эту схему, выделив принцип непричинения вреда в самостоятельную и равнозначную категорию, что позволило с большей чёткостью артикулировать моральные обязательства профессионала.

Таким образом, принципиализм Бичампа и Чайлдресса сформировался в результате интеграции ключевых положений деонтологии, утилитаризма и теории справедливости Ролза, а также их адаптации к конкретным потребностям биомедицинской практики через призму «Отчёта Белмонта». Эта синтезированная четырёхпринципная модель стала доминирующей нормативно-аналитической рамкой для разрешения этических дилемм в глобальном биоэтическом дискурсе, а именно обсуждении проблем, связанных с развитием медико-биологических наук и использованием медицинских технологий, с учётом этических ценностей и моральных норм.

ВКЛАД РОССИЙСКИХ ИССЛЕДОВАТЕЛЕЙ

Формирование и развитие биоэтического дискурса в России, обладающей собственной глубокой гуманистической и философско-религиозной традицией осмысления жизни и здоровья, внесли значимый вклад в критическое осмысление и адаптацию принципиалистской модели. Среди ключевых фигур, определивших специфику отечественного подхода, особо выделяются Б.Г. Юдин и П.Д. Тищенко.

Работы Б.Г. Юдина (1943–2021), академика РАН, одного из основоположников биоэтики в России, были направлены на преодоление узко-прикладного, рецептурного понимания принципов. Б.Г. Юдин подчёркивал необходимость их укоренения в более широком антропологическом и философском контексте. Он рассматривал биоэтику не просто как свод правил, а как форму защиты человека в условиях мощного технологического давления, особенно со стороны новых биомедицинских технологий.

В его интерпретации принципы автономии и благодеяния должны быть дополнены принципом «не навреди» в его экзистенциальном измерении, то есть как защитой целостности, идентичности и человеческого достоинства личности перед лицом манипулятивных возможностей науки. Б.Г. Юдин активно развивал идею о том, что созданные органы — этические комитеты — должны быть не просто контролирующими инстанциями, но площадками для глубокого междисциплинарного диалога [11].

П.Д. Тищенко, известный своей работой в области философии биомедицины и антропологии, внёс существенный вклад в критику абсолютизации принципа автономии в его либерально-индивидуалистической трактовке. Опираясь на традиции отечественной мысли, он аргументировал необходимость смещения акцента в сторону коммуникативной и диалогической природы принятия решений в медицине. Его концепция «коммуникативного сообщества» предполагает, что этический выбор рождается не в изолированном сознании автономного индивида, а в пространстве диалога между пациентом, врачом, семьёй и социокультурным контекстом. Это позволяет критически переосмыслить западную модель информированного согласия, дополнив её аспектом совместного понимания и разделённой ответственности [11]. Для того, чтобы оценить и понять, какие конкретно возникают проблемы с трансформацией или изменением биоэтических подходов, необходимо рассмотреть их на конкретных примерах.

ПРИНЦИП БЛАГОДЕЯНИЯ

В области биоэтики принципы благодеяния («твори добро») и «не навреди» (лат. *primum non nocere*) представляют собой два взаимодополняющих этических постулата, которые обязывают медицинских работников и исследователей действовать в интересах пациентов и общества, стремясь предотвратить или минимизировать потенциальный вред. Принцип «не навреди» предписывает избегать действий, способных причинить ущерб, тогда как принцип благодеяния требует активного стремления к улучшению здоровья и благосостояния, выходящего за рамки простого воздержания от причинения вреда. Данный принцип в современных условиях сталкивается с определёнными вызовами, а именно с такими понятиями, как «чёрный ящик» и клиническая валидация.

ИИ обладает возможностью обучения/самообучения. Для этого используются следующие подходы.

1. Глубокое обучение и нейронные сети. ИИ обучается на больших наборах данных, чтобы распознавать и классифицировать патологии. Например, свёрточные нейронные сети.

2. Машинное обучение. Алгоритмы совершенствуются (обучаются на данных) и делают прогнозы или решения на их основе. В медицине машинное обучение может использоваться для классификации изображений, обнаружения аномалий и предсказания исходов лечения [12].

3. Методы обучения и обработки данных. Для успешного обучения моделей ИИ требуется большое количество данных, но в медицине сбор и разметка таких объёмов информации сопряжены с трудностями. Для преодоления этого используют трансферное обучение, аугментацию данных и генеративно-состязательные сети [13].

Область здравоохранения — один из самых впечатляющих примеров использования ИИ. К помощи искусственного интеллекта прибегают для диагностики заболеваний, анализа медицинских изображений (рентгеновские снимки, магнитно-резонансные томограммы, компьютерные томограммы), прогнозирования развития болезней и организации лечения. Например, алгоритмы глубокого обучения могут анализировать магнитно-резонансные томограммы или рентгеновские изображения для выявления признаков рака, заболеваний сердца или неврологических расстройств. В результате ускоряется процесс диагностики, снижается количество ошибок, связанных с человеческим фактором, и увеличивается точность диагностики, так как ИИ лучше, чем врач лучевой диагностики распознаёт норму и патологию, поскольку не устаёт, не отвлекается и др. [14–16]. Выявление переломов на рентгенограммах, объёмных образований в лёгком или молочной железе (профилактические осмотры), помощь при постановке диагноза болезни Альцгеймера, оценка величины поражения мозга при геморрагическом инсульте, контроль динамики лечебного процесса — области, где ИИ очень полезен.

Современные системы ИИ помогают врачам в выборе наиболее эффективного метода лечения на основе многочисленных данных о пациенте, используются для анализа генетической информации, что позволяет разрабатывать персонифицированные подходы к лечению и прогнозировать реакции на различные препараты. Так, система IBM Watson в области медицины активно применяется для помощи врачам в принятии решений по лечению рака и других сложных заболеваний [14, 17]. Такие области медицины, как онкология, неврология, кардиология [17, 18] сегодня активно используют ИИ-помощника. Однако необходимо понимать, что ИИ обучается на конкретных примерах, и если они не достаточны корректны, то он тиражирует неточности и ошибки при анализе новых изображений [18].

Алгоритм для диагностики пневмонии по рентгеновским снимкам, разработанный в одной больнице, показал высочайшую точность при применении на её территории. Однако при внедрении этого алгоритма в другую больницу его точность резко упала. Оказалось, ИИ научился распознавать не патологические изменения в лёгких, а метаданные: специфический артефакт, который оставлял конкретный рентген-аппарат в первом госпитале. Врачи не могли этого знать, поскольку не понимали логики работы алгоритма. Другими словами, возникает проблема невозможности провести независимую клиническую экспертизу решения алгоритма. Врач слепо доверяет результату, что может привести к фатальной ошибке.

DOI: https://doi.org/10.17816/KMJ690475

Другой вызов приводит к размытию ответственности врача. Это становится очевидным в ситуациях применения роботизированной медицины. Например, хирургическая система da Vinci выполняла операцию, управляемая хирургом. Алгоритм системы, обеспечивающий «дрожание» руки, дал сбой, приведя к повреждению сосуда. Кто виноват? Хирург, не остановивший операцию? Производитель оборудования, выпустивший некачественное программное обеспечение? Больница, не обеспечившая должное техобслуживание прибора? [19, 20]. Таким образом, цепочка ответственности становится чрезвычайно длинной и сложной, что затрудняет компенсацию ущерба пациенту.

Приведённые примеры ярко демонстрируют новые вызовы и указывают на необходимость оперативных мер по защите всех участников лечебно-диагностического процесса — как пациента, так и врача. Эти меры должны быть направлены на выполнение следующих требований:

• Обеспечение объяснимости ИИ (XAI). Разработка методов, визуализирующих принятие решений алгоритмом (например, тепловые карты, показывающие, на какие области снимка AI обратил внимание). Это позволит врачу не слепо доверять ИИ в постановке диагноза, а использовать результат анализа как дополнительный аргумент при формировании собственного клинического заключения.

• Подтверждение ключевой роли врача в принятии медицинских решений. Обязательное условие, чтобы окончательное клиническое решение принимал врач, а ИИ выступал лишь в роли ассистента, предоставляющего рекомендации.

• Нормативное закрепление ответственности. Разработка чётких юридических протоколов, определяющих долю вины разработчика, врача или учреждения в зависимости от характера и причин ошибки.

ПРИНЦИП УВАЖЕНИЯ АВТОНОМИИ ЛИЧНОСТИ

Принцип уважения автономии личности в биоэтике представляет собой ключевое этическое требование, которое подразумевает необходимость признания и уважения права каждого индивида на самоопределение и принятие собственных решений в вопросах, касающихся его здоровья, жизни и благополучия. Это включает право на информированное согласие при проведении медицинских процедур. Согласие пациента на то или иное медицинское вмешательство — сегодня реалии времени, стандарт медицины. Без согласия пациента (или в ряде случаев его представителя) врач не может/не имеет право на осмотр, проведение диагностических, лечебных, реабилитационных процедур. Данный принцип также подвергается определённым воздействиям со стороны цифровых технологий.

Одно из таких изменений — необходимость переосмысления информированного согласия в условиях «обучающихся» систем. Например, пациент подписывает согласие на использование ИИ-системы для мониторинга ЭКГ. Однако алгоритм постоянно обучается на новых данных, и через год его логика принятия решений может кардинально измениться. На какой именно алгоритм пациент давал согласие изначально? Здесь возникает проблема, когда традиционное согласие статично, в то время как ИИ-системы динамичны и эволюционируют.

В других ситуациях возникает необходимость обеспечить право пациента на объяснение. Например, алгоритм, анализирующий магнитно-резонансную томограмму, рекомендует отказать пациенту в сложной операции, оценив шансы на успех как крайне низкие. Пациент спрашивает: «Почему?». Врач может только сказать: «Так посчитала программа». Это нарушает право пациента на понимание факторов, влияющих на его лечение и судьбу, что вызывает проблему, когда отсутствие объяснения подрывает автономию пациента и его способность участвовать в принятии решений.

Представляется, что можно предложить следующие варианты решения:

• Многоуровневое и динамическое согласие. Внедрение цифровых платформ, где пациент может в любой момент получить информацию о версии алгоритма, на каких данных он обучался, а также подтвердить или отозвать своё согласие на его применение.

• Закрепление «права на объяснение» на законодательном уровне, как это сделано в GDPR (General Data Protection Regulation — общий регламент по защите данных) в ЕС. Это постановление Европейского союза, которое определяет правила сбора, обработки, хранения и распространения персональных данных на территории ЕС, обязывающее разработчиков создавать интерфейсы, способные генерировать объяснимые выводы для врача и пациента [21, 22].

ПРИНЦИП СПРАВЕДЛИВОСТИ

Принцип справедливости в области биоэтики предполагает равное и справедливое распределение преимуществ и рисков, связанных с научными исследованиями, медицинскими услугами и доступом к ресурсам здравоохранения, среди всех слоёв населения, вне зависимости от их социального, экономического или другого положения. Этот принцип требует равного обращения к тем, кто находится в одинаковых условиях, и неравного — к тем, кто находится в различных обстоятельствах, а также акцентирует внимание на необходимости устранения препятствий, которые ограничивают доступ к медицинским услугам. В контексте внедрения ИИ в медицину этот принцип также подвергается различным вызовам. Среди них можно назвать алгоритмическую предвзятость. Например, широко известный пример с алгоритмом Optum, который использовался в США для управления ресурсами здравоохранения. Алгоритм присваивал одинаковый уровень риска чернокожим и белым пациентам только в том случае, если состояние чернокожих пациентов

было значительно тяжелее. Причина заключалась в том, что модель обучалась на исторических данных о затратах на здравоохранение. Из-за системного неравенства доступ к медпомощи для чернокожих пациентов исторически был ограничен, что приводило к более низким затратам на их лечение. Алгоритм интерпретировал это как «меньшую нуждаемость в помощи» [23].

Можно привести другой пример, когда коммерческие алгоритмы для обработки изображений кожи (дерматоскопия) показывают значительно худшую точность диагностики меланомы для тёмных типов кожи, так как они обучались преимущественно на изображениях светлой кожи. Всё это приводит к тому, что ИИ не устраняет предрассудки, а, напротив, кодифицирует и масштабирует их, формируя систему, способную дискриминировать уязвимые группы.

Возможны следующие пути решения:

• Обязательный аудит на справедливость: перед внедрением любой медицинской AI-системы должен проводиться независимый аудит на разнообразных датасетах, репрезентирующих все группы населения.

• Обеспечение разнообразия данных: стимулирование создания открытых, аннотированных и разнообразных медицинских датасетов.

• Прозрачность: публикация разработчиками информации о демографическом составе данных для обучения и тестирования.

В контексте биоэтики конфиденциальность представляет собой обязательство медицинских работников и других специалистов сохранять в тайне информацию о пациентах, полученную в ходе оказания медицинской помощи. Это необходимо для защиты их личной жизни, социального статуса и экономических интересов.

ПРИНЦИП КОНФИДЕНЦИАЛЬНОСТИ

Принцип конфиденциальности играет важную роль в установлении доверительных отношений между врачами и пациентами, что способствует откровенности пациентов и предотвращает потенциальную стигматизацию и дискриминацию, связанные с их медицинскими данными. Данный принцип тесно связан с Клятвой Гиппократа, которая подчёркивает важность уважения к пациенту и сохранения его тайны как основополагающего элемента этической практики в медицине.

Это является серьёзным вызовом. Для обучения мощных алгоритмов требуются огромные массивы персональных медицинских данных пациентов. Их централизация в одном месте создаёт колоссальный риск утечки информации [24, 25].

Возможные пути решения включают следующее.

• Объединённое обучение: передовая техника, при которой алгоритм обучается децентрализованно. Данные не покидают серверы больницы, на них локально обучается модель, и на центральный сервер отправляются только обновления параметров модели, а не сами данные. Это резко снижает риски для конфиденциальности.

• Дифференциальная приватность: добавление в данные статистического «шума», который делает невозможным идентификацию конкретного человека, но сохраняет общие статистические закономерности для обучения алгоритмов.

Данный анализ показывает, что биоэтика стоит на пороге парадигмальных изменений. Мы движемся от этики человека к этике систем, что требует определённых действий:

1. Расширение этической границы: принципы Бичампа («не навреди», «делай благо», справедливость и уважение автономии личности) должны быть дополнены принципами, отражающими технологическую реальность:

• прозрачность — стремление к максимальной объяснимости алгоритмических решений;

• подотчётность — чёткое законодательное определение ответственности за каждое звено в цепочке «разработчик — врач — учреждение»;

• справедливость — техническая и социальная справедливость, заложенная в дизайн алгоритма.

2. Изменение образования: врачи будущего должны обладать не только медицинской, но и алгоритмической грамотностью — понимать основы работы ИИ, его ограничения и потенциальные ошибки [25].

3. Принцип опережающего регулирования: регуляторы (например, Росздравнадзор в условиях Российской Федерации) должны разрабатывать стандарты и проводить валидацию медицинских AI-систем не только на эффективность, но и на этическую корректность.

ИИ не просто новый инструмент в медицине, а агент трансформации самой этической ткани здравоохранения. Переосмысление биоэтических принципов — это не отмена прошлого, а их важная и необходимая эволюция. Успех интеграции ИИ в медицину будет измеряться не только его точностью и эффективностью, но и его способностью укреплять, а не подрывать, фундаментальные ценности медицины: доверие, справедливость и уважение к автономии пациента. Достижение этой цели требует консолидированных усилий разработчиков, врачей, биоэтиков, юристов и регуляторов[5] [26, 27].

Регулировать искусственный интеллект — сложная и дуальная задача. Чрезмерное регулирование ИИ может отрицательно сказаться на инновациях, в то время

[5] ResearchGate [интернет]. 2008–2025. ResearchGate GmbH. Режим доступа: https://www.researchgate.net/?ref=logo&_sg=Ro8tDhb7EvgtUnJVWpW2kmsOvwHW2LdsHfPvsp104ts4jDuycM63CeWrDnPF_iVAGuoP4KPMLNeGbIY&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uliwicGFnZSI6InB1YmxpY2F0aW9uliwicG9zaXRpb24iOiJnbG9iYWxlZWFkZXIifX0 Дата обращения: 17.09.2025.

как недостаточное регулирование ИИ может привести к серьёзному ущербу прав граждан, а также к потере возможности формировать будущее цивилизованное общество [26].

Основной задачей ИИ является автоматизированное исполнение некоторых однотипных задач, когда использование машинного обучения (Machine Learning) и глубокого обучения (Deep Learning) необходимо для компьютерного обучения [28–30].

ЗАКЛЮЧЕНИЕ

Выявлены фундаментальные противоречия между классическими принципами медицинской этики и реалиями современной медицины при использовании технологии искусственного интеллекта — это столкновение принципа благодеяния с проблемой «чёрного ящика» и размывания ответственности, несовершенство принципа автономии и необходимость пересмотра модели информированного согласия для динамических алгоритмов, а также ущербность принципов справедливости и конфиденциальности.

Сохранение доверия в медицине требует не отказа от традиционных принципов, а их эволюции. Нужно дополнить их новыми столпами: прозрачностью, подотчётностью и технической справедливостью. Это предполагает разработку новых регуляторных стандартов, обязательный аудит алгоритмов и интеграцию этического дизайна в процесс разработки медицинского ИИ.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Вклад авторов. Н.Ф.Т. — определение концепции, пересмотр и редактирование рукописи, руководство исследованием; Г.М.Э. — проведение исследования, написание черновика рукописи, пересмотр и редактирование рукописи. Все авторы одобрили рукопись, а также согласились нести ответственность за все аспекты работы, гарантируя надлежащее рассмотрение и решение вопросов, связанных с точностью и добросовестностью любой её части.

Источники финансирования. Отсутствуют.

Раскрытие интересов. Авторы заявляют об отсутствии отношений, деятельности и интересов за последние три года, связанных с третьими лицами (коммерческими и некоммерческими), интересы которых могут быть затронуты содержанием статьи.

Оригинальность. При создании настоящей работы авторы не использовали ранее опубликованные сведения (текст, иллюстрации, данные).

Доступ к данным. Редакционная политика в отношении совместного использования данных к настоящей работе не применима, новые данные не собирали и не создавали.

Генеративный искусственный интеллект. При создании настоящей статьи технологии генеративного искусственного интеллекта не использовали.

Рассмотрение и рецензирование. Настоящая работа подана в журнал в инициативном порядке и рассмотрена по обычной процедуре. В рецензировании участвовали три внешних рецензента, член редакционной коллегии и научный редактор издания.

ADDITIONAL INFORMATION

Author contributions: N.F.T.: conceptualization, supervision, writing—review & editing; G.M.E.: investigation, writing—original draft, writing—review & editing. All the authors approved the version of the manuscript to be published and agreed to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding sources: No funding.

Disclosure of interests: The authors have no relationships, activities, or interests for the last three years related to for-profit or not-for-profit third parties whose interests may be affected by the content of the article.

Statement of originality: No previously obtained or published material (text, images, or data) was used in this study or article.

Data availability statement: The editorial policy regarding data sharing does not apply to this work, as no new data was collected or created.

Generative AI: No generative artificial intelligence technologies were used to prepare this article.

Provenance and peer-review: This paper was submitted unsolicited and reviewed following the standard procedure. The peer review process involved three external reviewers, a member of the Editorial Board, and the in-house science editor.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

1. Averkin AN, Afanasev SD, Mikryukov AA, et al. Big data standardization: international and national standards. *Information society*. 2021;(4–5):220–258. doi: 10.52605/16059921_2021_04_220 EDN: DPGURW

2. Garbuk SV, Shalaev AP. Prospective structure of national standards in the field of artificial intelligence. *Standards and quality*. 2021;(10):26–33. doi: 10.35400/0038-9692-2021-10-26-33 EDN: IEOWVB

3. *Ethics and Digital: From Problem to Solution*. Potapov EG, Shklyaruk MS, editors. Moscow: Russian Presidential Academy of National Economy and Public Administration. 2021. 184 p. (In Russ.)

4. Nikiforov SV. Legal regulation and formalization of the artificial intelligence's legal personality in Russian and international law. *Gaps in Russian legislation*. 2020;(1):79–82. EDN: IXZITB

5. Mindigulova AA. Ethics and artificial intelligence: problems and contradictions. *Meditsina. Sotsiologiya. Filosofiya. Prikladnye issledovaniya*. 2022;(3):146–150. EDN: MENBGM

6. Mittelstadt B. Principles alone cannot guarantee ethical AI. *Nature machine intelligence*. 2019;1(11):501–507. doi: 10.1038/s42256-019-0114-4 EDN: OAJKFF

7. Beauchamp TL, James F. *Childress Principles of Biomedical Ethics*. Oxford University Press, 2001. 454 p.

8. Nezhmetdinova FT. Bioethics in the context of contemporary scientific strategies and applied ethics in the age of contemporary technologies. *Vestnik of Saint Petersburg University. International Relations*. 2009;(1):221–229. EDN: KWLOJN

9. Department of Health, Education, and Welfare; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. *J Am Coll Dent*. 2014;81(3):4–13. Available from: https://pubmed.ncbi.nlm.nih.gov/25951677/ Accessed: 01.05.2025.

10. Rawls D. *Theory of justice*. Cambridge, Massachusetts: Belknap Press of Harvard University Press; 1971. 538 p.

11. Nezhmetdinova FT, Guryleva ME. Russian school of bioethics: a quarter of a century of development. *Kazan medical journal*. 2018;99(3):521–527. doi: 10.17816/KMJ2018-521 EDN: XNKIRV

12. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care—Addressing Ethical Challenges. *N Engl J Med*. 2018;378(11):981–983. doi: 10.1056/NEJMp1714229

13. Ting Sim JZ, Fong QW, Huang W, Tan CH. Machine learning in medicine: what clinicians should know. *Singapore Med J*. 2023;64(2):91–97. doi: 10.11622/smedj.2021054 EDN: UKWAOM

14. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500–510. doi: 10.1038/s41568-018-0016-5
15. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab Syndr*. 2020;14(4):337–339. doi: 10.1016/j.dsx.2020.04.012 EDN: EUGDAX
16. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230–243. doi: 10.1136/svn-2017-000101
17. Korabelnikov DI, Lamotkin AI. The effectiveness of using artificial intelligence in clinical medicine. *Farmakoekonomika. Modern pharmacoeconomics and pharmacoepidemiology*. 2025;18(1):114–124. doi: 10.17749/2070-4909/farmakoekonomika.2025.287 EDN: OHRVVR
18. Chen J, See KC. Artificial Intelligence for COVID-19: Rapid Review. *J Med Internet Res*. 2020;22(10):e21476. doi: 10.2196/21476
19. Vvedenskaya EV. Transformation of the physician-patient relationship: from bioethics to roboethics. *Human being*. 2023;34(6):65–83. doi: 10.31857/S023620070029305-2 EDN: LQEJGB
20. Tsomartova FV. Robotization in healthcare: legal perspective. *Health care of the Russian Federation*. 2020;64(2):88–96. doi: 10.46563/0044-197X-2020-64-2-88-96 EDN: ENSOKC
21. Price WN 2nd, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. *JAMA*. 2019;322(18):1765–1766. doi: 10.1001/jama.2019.15064
22. *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization; 2021. 148 p. License: CC BY-NC-SA 3.0 IGO ISBN: 978-92-4-002920-0
23. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. doi: 10.1126/science.aax2342
24. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. *PLoS Med*. 2018;15(11):e1002689. doi: 10.1371/journal.pmed.1002689
25. Morley J, Elhalal A, Garcia F, et al. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds Mach*. 2021;31(2):239–256. doi: 10.1007/s11023-021-09563-w EDN: QYBMRI
26. Bryzgalina EV, Gumarova AN, Shkomova EM. Key problems, risks and restrictions of using artificial intelligence in medicine and education. *Moscow university bulletin. Series 7. Philosophy*. 2022;(6):93–108. EDN: GXUYWB
27. Nezhmetdinova FT, Guryleva ME, Blatt NL. New Role of Bioethics in Emergency Situations on the Example of COVID-19. *Bionanoscience*. 2022;12(2):620–626. doi: 10.1007/s12668-021-00915-5 EDN: UQONAY
28. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–1930. doi: 10.1161/CIRCULATIONAHA.115.001593 EDN: VFLTDF
29. Rabbani N, Kim GYE, Suarez CJ, Chen JH. Applications of machine learning in routine laboratory medicine: Current state and future directions. *Clin Biochem*. 2022;103:1–7. doi: 10.1016/j.clinbiochem.2022.02.011 EDN: TUMKNG
30. Choi RY, Coyner AS, Kalpathy-Cramer J, et al. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol*. 2020;9(2):14. doi: 10.1167/tvst.9.2.14

ОБ АВТОРАХ

*** Нежметдинова Фарида Тансыковна**, канд. филос. наук, доцент;
адрес: Россия, 420015, Казань, ул. К. Маркса, д. 65;
ORCID: 0000-0003-2875-128X;
eLibrary SPIN: 8441-6943;
e-mail: nadgmi@mail.ru

Гурылева Марина Элисовна, д-р мед. наук, профессор;
ORCID: 0000-0003-2772-129X;
eLibrary SPIN: 6207-9971;
e-mail: meg4478@mail.ru

AUTHORS INFO

* **Farida T. Nezhmetdinova**, Cand. Sci. (Philosophy), Assistant Professor;
address: 65 K. Marx st, Kazan, Russia, 420015;
ORCID: 0000-0003-2875-128X;
eLibrary SPIN: 8441-6943;
e-mail: nadgmi@mail.ru

Marina E. Guryleva, MD, Dr. Sci. (Medicine), Professor;
ORCID: 0000-0003-2772-129X;
eLibrary SPIN: 6207-9971;
e-mail: meg4478@mail.ru

* Автор, ответственный за переписку / Corresponding author