

УДК 004.8

DOI: [10.26102/2310-6018/2025.49.2.016](https://doi.org/10.26102/2310-6018/2025.49.2.016)

Оценка надежности и эффективности систем искусственного интеллекта в лучевой диагностике на этапе эксплуатации

В.В. Зинченко¹, А.В. Владзимирский¹, К.М. Арзамасов^{1,2}

¹Научно-практический клинический центр диагностики и телемедицинских технологий
Департамента здравоохранения города Москвы, Москва, Российская Федерация

²МИРЭА – Российский технологический университет, Москва, Российская Федерация

Резюме. В условиях активного внедрения технологий искусственного интеллекта (ИИ) в здравоохранение особую актуальность приобретает обеспечение стабильного, контролируемого и высокого качества работы таких систем на этапе эксплуатации. Мониторинг систем ИИ закреплён на законодательном уровне: в течение трех лет после внедрения медицинских изделий, включая системы ИИ, необходимо предоставлять регулярные отчеты в контролирующие органы. Целью исследования является разработка методов оценки надежности и эффективности медицинского искусственного интеллекта для лучевой диагностики. Апробация предложенных методов проведена на данных Московского Эксперимента по использованию инновационных технологий в области компьютерного зрения по направлению рентгенологии органов грудной клетки, собранных за 2023 год. Разработанные методы учитывают комплекс параметров: возникающие технологические дефекты, время обработки исследований, степень согласия врачей с результатами анализа и другие показатели. Предложенный подход может быть адаптирован для различных видов медицинских исследований и стать основой для комплексной оценки систем ИИ в рамках мониторинга медицинских изделий с искусственным интеллектом. Внедрение данных методов способно повысить уровень доверия медицинского сообщества не только к конкретным решениям на базе ИИ, но и к интеллектуальным технологиям в здравоохранении в целом.

Ключевые слова: искусственный интеллект, надежность, эффективность, система искусственного интеллекта, рентгенология, лучевая диагностика, мониторинг.

Благодарности: Статья подготовлена авторским коллективом в рамках НИР «Научно-методические основы цифровой трансформации службы лучевой диагностики», (№ ЕГИСУ: № 123031400118-0) в соответствии с Приказом от 21.12.2022 г. № 1196 «Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы государственным бюджетным (автономным) учреждениям подведомственным Департаменту здравоохранения города Москвы, на 2023 год и плановый период 2024 и 2025 годов» Департамента здравоохранения города Москвы.

Для цитирования: Зинченко В.В., Владзимирский А.В., Арзамасов К.М. Оценка надежности и эффективности систем искусственного интеллекта в лучевой диагностике на этапе эксплуатации. *Моделирование, оптимизация и информационные технологии.* 2025;13(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1886> DOI: 10.26102/2310-6018/2025.49.2.016

Assessment of the reliability and effectiveness of artificial intelligence systems in radiation diagnostics at the operational stage

V.V. Zinchenko¹, A.V. Vladzimirskyy¹, K.M. Arzamasov^{1,2}

¹Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, Moscow, the Russian Federation

²MIREA – Russian Technological University, Moscow, the Russian Federation

Abstract. In the context of the active implementation of artificial intelligence (AI) technologies in healthcare, ensuring stable, controlled and high-quality operation of such systems at the operational stage is of particular relevance. Monitoring of AI systems is enshrined in law: within three years after the implementation of medical devices, including AI systems, it is necessary to provide regular reports to regulatory authorities. The aim of the study is to develop methods for assessing the reliability and effectiveness of medical artificial intelligence for radiation diagnostics. The proposed methods were tested on the data of the Moscow Experiment on the use of innovative technologies in the field of computer vision in the direction of chest radiography, collected in 2023. The developed methods take into account a set of parameters: emerging technological defects, research processing time, the degree of agreement of doctors with the analysis results and other indicators. The proposed approach can be adapted for various types of medical research and become the basis for a comprehensive assessment of AI systems as part of the monitoring of medical devices with artificial intelligence. The implementation of these methods can increase the level of trust of the medical community not only in specific AI-based solutions, but also in intelligent technologies in healthcare in general.

Keywords: artificial intelligence, reliability, efficiency, artificial intelligence system, radiology, radiation diagnostics, monitoring.

Acknowledgments: The article was prepared by a team of authors within the framework of the research work "Scientific and methodological foundations of digital transformation of the radiation diagnostic service" (EGISU No.: No. 123031400118-0) in accordance with the Order of 21.12.2022 No. 1196 "On approval of state assignments, financial support for which is carried out at the expense of the Moscow city budget to state budgetary (autonomous) institutions subordinate to the Moscow City Department of Health, for 2023 and the planning period of 2024 and 2025" of the Moscow City Department of Health.

For citation: Zinchenko V.V., Vladzimirskyy A.V., Arzamasov K.M. Assessment of the reliability and effectiveness of artificial intelligence systems in radiation diagnostics at the operational stage. *Modeling, Optimization and Information Technology*. 2025;13(2). (In Russ.). URL: <https://moitvvt.ru/journal/pdf?id=1886> DOI: 10.26102/2310-6018/2025.49.2.016

Введение

По мере того, как создание и практическое применение систем искусственного интеллекта (СИИ) активно развивается в последние годы, задачи обеспечения устойчивого уровня безопасности и качества работы СИИ становятся все более важными. В этой связи требуются дополнительные меры, в первую очередь научного и методического характера, для контроля и обеспечения, в том числе надежности, эффективности работы решений на основе искусственного интеллекта [1, 2]. Нестабильность метрик точности моделей машинного обучения при их эксплуатации в условиях реальной клинической практики часто недооценивается или даже игнорируется при создании и применении СИИ. Недавние публикации по этой теме показывают, что процедуры оценки и контроля на пострегистрационном этапе жизненного цикла СИИ требуют существенных улучшений¹ [3].

Причинами данной потребности является то, что во время эксплуатации СИИ нередко ухудшаются/изменяются некоторые показатели работоспособности, продемонстрированные на этапе независимых клинических испытаний в целях государственной регистрации и вывода на рынок. Одна из причин такого изменения состоит в том, что во время эксплуатации, в СИИ поступают данные, которые могли отсутствовать в обучающих, тестовых или валидационных наборах данных. Известная как «концептуальный дрейф», изменчивость данных неизбежно влияет на качество работы моделей машинного обучения, вплоть до того, что модель может больше

¹ Artificial Intelligence and Machine Learning in Software as a Medical Device. Food & Drug Administration. URL: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> (дата обращения: 02.03.2025).

работать неадекватно новой реальности, т. е. генерировать больший уровень ошибок в определенной медицинской организации, чем это было заявлено на этапе разработки [4].

Таким образом, с точки зрения эффективности и надежности работы СИИ, риск ухудшения работы моделей машинного обучения представляет собой новую и актуальную проблему.

Современные научные исследования и систематические обзоры демонстрируют необходимость всесторонней оценки СИИ, включающей как технические, так и клинические аспекты в процессе их функционирования. Исследование Pemberton и соавторов указывает на существенный дисбаланс: технический мониторинг СИИ осуществляется значительно чаще клинического [5]. Примечательно, что даже при оценке СИИ одинакового профиля разработчики применяют различные методологические подходы. Аналогичные выводы о преобладании технического мониторинга над клиническим представлены в работе Nomura и соавторов, где предлагается инновационная методика разработки, обучения и валидации программного обеспечения (ПО) с акцентом на технические характеристики [6].

На основании вышеизложенного, целью настоящего исследования является разработка методологии оценки эффективности и надежности СИИ, применяемых в радиологической службе. Методология базируется на комплексном анализе различных показателей функционирования СИИ и последующем формировании обоснованных заключений о целесообразности их дальнейшего внедрения в практику здравоохранения. Для реализации поставленной цели проводился всесторонний анализ использования СИИ в медицинской сфере с учетом как технических, так и клинических аспектов. Разрабатываемые методы проектировались с перспективой их интеграции в качестве ключевого элемента системы комплексной оценки СИИ в рамках постоянного мониторинга их функционирования.

Материалы и методы

Для тестирования разработанных методов использовались данные, полученные в рамках Московского эксперимента по внедрению технологий компьютерного зрения для анализа медицинских изображений (<https://mosmed.ai/>). На основе доступной информации были сформированы выборки рентгенографических исследований грудной клетки (РГ), проведенных в 2023 году в течение 10 последовательных месяцев. В данном виде исследований на момент сбора данных функционировало пять различных СИИ от разных разработчиков (Таблица 1).

В процессе анализа из сформированных выборок были извлечены разнообразные параметры, предусмотренные разрабатываемыми методиками оценки надежности и эффективности СИИ. В частности, были собраны данные о технологических дефектах, длительности обработки исследований (Т), показателях площади под ROC-кривой (CAUC) и других характеристиках, отражающих функциональность систем.

Для анализа технологических дефектов использовались временные метрики обработки исследований СИИ и классификация ошибок, препятствующих получению результатов. Дополнительно применялись данные, предоставленные оператором Единого радиологического информационного сервиса (ЕРИС), который внедрил автоматические алгоритмы выявления различных технологических дефектов. Оценка показателя площади под ROC-кривой (AUC/CAUC) проводилась путем сопоставления результатов работы СИИ с текстовыми протоколами врачей-рентгенологов и вероятностными оценками наличия патологий. Для определения уровня согласованности с врачебными заключениями и идентификации специфических дефектов (типа г, д) квалифицированные врачи-рентгенологи ежемесячно анализировали выборку из 80 рентгенографических исследований. Данный объем выборки, несмотря на ограничения Московского Эксперимента по привлечению экспертов, соответствует

требованиям репрезентативности (сохраняя необходимое соотношение нормы и патологии согласно базовым диагностическим требованиям к СИИ), учитывает риски для производителей и потребителей, а также технические спецификации СИИ [7]. За весь период наблюдения (10 отчетных периодов) было проанализировано 1484855 исследований для выявления определенных типов дефектов (б, в) и оценки эффективности, за исключением анализа согласованности с врачебными заключениями.

Таблица 1 – Информация о СИИ, использованных в апробации методологии оценки надежности и эффективности в реальной клинической практике

Table 1 – Information on artificial intelligence systems used in testing methods for assessing reliability and effectiveness in real clinical practice

№	ПО с ИИ	Область применения	Назначение по информации от разработчика
1	СИИ_1	РГ ОГК	Анализ РГ изображений ОГК в прямой проекции на наличие признаков злокачественных новообразований, туберкулеза, профессиональных болезней и других скрыто протекающих заболеваний ОГК.
2	СИИ_2	РГ ОГК	Результат работы отражает наличие хотя бы одного рентгенологического признака из перечня: плевральный выпот, пневмоторакс, ателектаз, очаг затемнения, инфильтрация/консолидация, диссеминация, полость с распадом, полость с уровнем жидкости, кальцинат / кальцинированная тень в легких, нарушение целостности кортикального слоя.
3	СИИ_3	РГ ОГК	СИИ выявляет на РГ ОГК и флюорограммах наличие диагностических признаков туберкулеза, пневмонии, гнойных и некротических состояний, образований в легком и патологий средостения, а также определяет гидроторакс, пневмоторакс, кардиомегалию и переломы ребер.
4	СИИ_4	РГ ОГК	СИИ выявление признаков различных заболеваний. Формирование полного проекта протокола, расчет кардиоторакального индекса, процента обрезки лёгких.
5	СИИ_5	РГ ОГК	СИИ автоматически оценивает вероятность наличия патологий на рентгенограммах ОГК и флюорограммах, осуществляет поиск и локализацию патологических признаков.

В ходе проведения исследования были разработаны методы оценки эффективности и надежности работы СИИ в условиях реальной клинической практики (Рисунок 1).

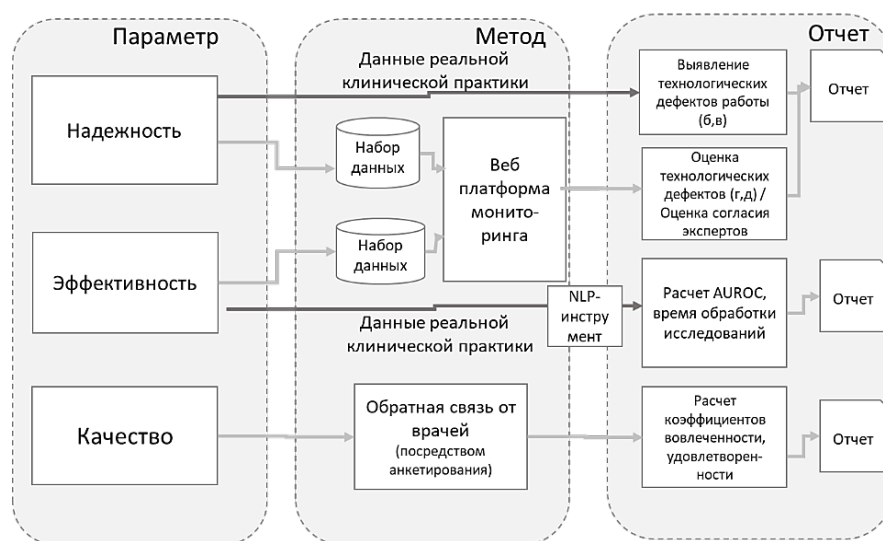


Рисунок 1 – Схема метода сбора и анализа данных реальной клинической практики
Figure 1 – The scheme of the method of collecting and analyzing data from real clinical practice

Предложенные методы оценки были внедрены в контур Московского Эксперимента в рамках проведения технологических и клинических мониторингов, на схеме на Рисунке 2 внедренные методы по оценке надежности и эффективности СИИ представлены в виде «Платформа комплексной оценки ПО», которая также дополнительно включает оценку качества медицинских СИИ.

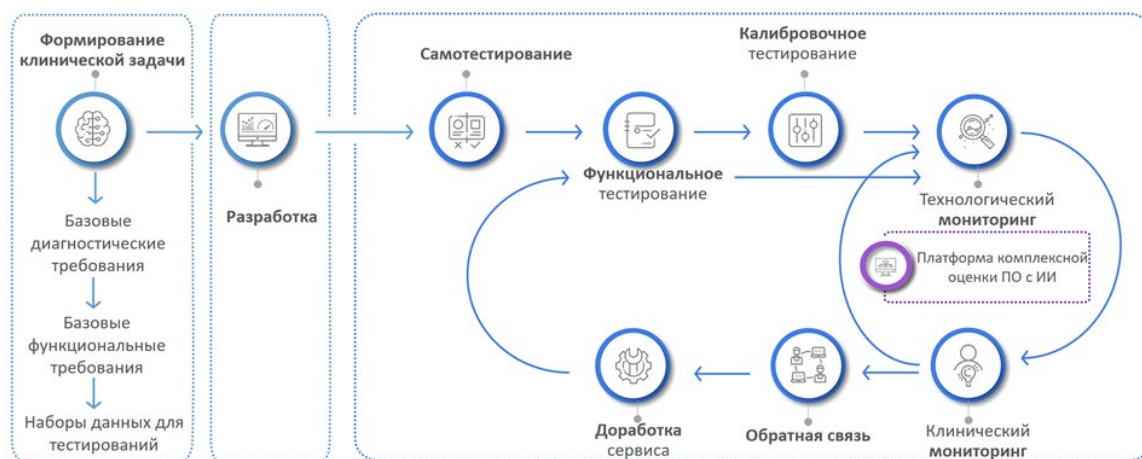


Рисунок 2 – Общая схема работы Эксперимента

Figure 2 – The general scheme of the Experiment

Технологические дефекты, представленные в данной работе, или параметры технической надежности, другими словами, способны оказывать влияние на безопасность пациентов, а именно – некорректная работа заявленного функционала может привести к ошибочному решению врача и, как следствие, причинить вред пациенту; при некорректном отображении области исследований могут быть пропущены патологии на изображении и, как следствие, также причинить потенциальный вред пациенту и т. п.

Для оценки надежности работы СИИ были выбраны технологические дефекты следующих видов (согласно номенклатуре Московского Эксперимента по внедрению компьютерного зрения):

- Тип «б»: отсутствие результатов проанализированных исследований, а также выявленные ошибки на стороне СИИ при возвращении выходных данных.
- Тип «в»: некорректная работа заявленного функционала СИИ, затрудняющая работу врача-рентгенолога или делающая ее выполнение невозможным с надлежащим качеством.
- Тип «г»: дефекты, связанные с отображением области изображений.
- Тип «д»: иные нарушения целостности и содержимого файлов с результатами исследований.

Расчет коэффициента надежности C_{td} был выполнен с использованием следующей формулы (1):

$$C_{td} = k_0 + k_{td,b} \cdot C_{td,b} + k_{td,v} \cdot C_{td,v} + k_{td,r,d} \cdot C_{td,r,d}, \quad (1)$$

$$C_{td,b} = \frac{B + E_{СИИ}}{N - E_{РАС}}, \quad (2)$$

$$C_{td,v} = \frac{V}{N - E_{РАС}}, \quad (3)$$

$$C_{td,r,d} = \frac{GD}{N_s}, \quad (4)$$

где $k_{td,b}$ – коэффициент взвешенности для дефектов типа «б»; В – количество обработанных с помощью СИИ исследований с технологическими дефектами типа «б»; $E_{сии}$ – количество обработанных с помощью СИИ исследований с ошибками на стороне СИИ; N – общий объем направленных данных за отчетный период; E_{PACS} – количество обработанных исследований с ошибками на стороне PACS (ЕРИС); $k_{td,b}$ – коэффициент взвешенности для дефектов типа «в»; V – количество обработанных с помощью СИИ исследований с технологическими дефектами типа «в»; $k_{td,g,d}$ – коэффициент взвешенности для дефектов типа «г, д»; GD – количество обработанных с помощью СИИ исследований с технологическими дефектами типа «г» и «д»; N_s – объем данных, пересмотренных экспертами ($N_s = 80$).



Рисунок 3 – Схема работы программы оценки коэффициентов надежности, эффективности и качества

Figure 3 – The scheme of the program for evaluating reliability, efficiency and quality coefficients

Для обеспечения автоматизированной оценки коэффициентов надежности и эффективности было разработано ПО², с использованием языка python (Рисунок 4), где модуль статистического анализа формирует статистику на основании тестов: ANOVA/теста Манна-Уитни для сравнения составляющих для разных СИИ с заданным уровнем значимости, например, $p < 0,05$: сравнение отдельных дефектов, параметров AUC, согласие с заключением врачей, согласие с локализацией, время отклика СИИ. Для мониторинга и оценки эффективности СИИ в клинической практике разработанная методология предусматривает верификацию соответствия времени обработки и диагностической точности (измеряемой площадью под характеристической кривой) установленным пороговым значениям. В комплексную оценку эффективности также была включена проверка согласованности между экспертными заключениями и результатами СИИ, как в текстовой форме, так и в отношении локализации патологических изменений на изображениях (Рисунок 5).

² Васильев Ю.А., Арзамасов К.М., Зинченко В.В. и др. Программа оценки коэффициентов безопасности, эффективности и качества при выполнении мониторинга системы искусственного интеллекта в лучевой диагностике: опубл. 17.04.2024. Свидетельство о государственной регистрации программы для ЭВМ № 2024618798.

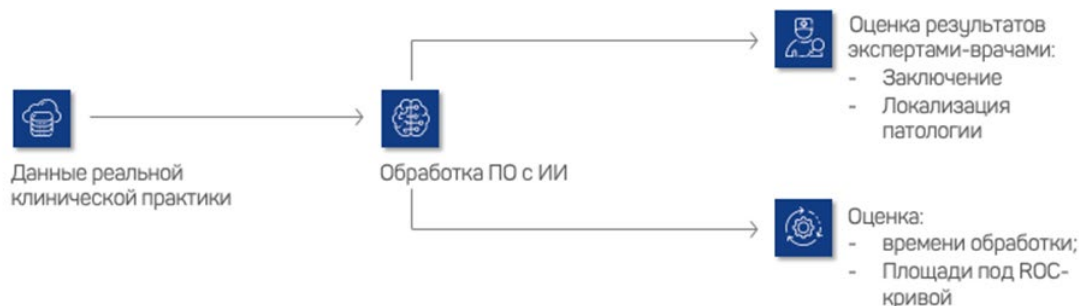


Рисунок 5 – Схема проведения мониторинга эффективности СИИ

Figure 5 – The scheme for monitoring the effectiveness of artificial intelligence system

Предельно допустимое время обработки было регламентировано в рамках Московского эксперимента и составило 6,5 минут. Оценка показателя AUC проводилась посредством анализа площади под ROC-кривой, построенной на основе выходных данных СИИ и результатов интерпретации протоколов врачей-рентгенологов по соответствующим исследованиям. Для этого применялся специализированный NLP-инструмент (система анализа текстовой информации на основе алгоритмов искусственного интеллекта), позволяющий классифицировать исследования по категориям «норма»/«патология».

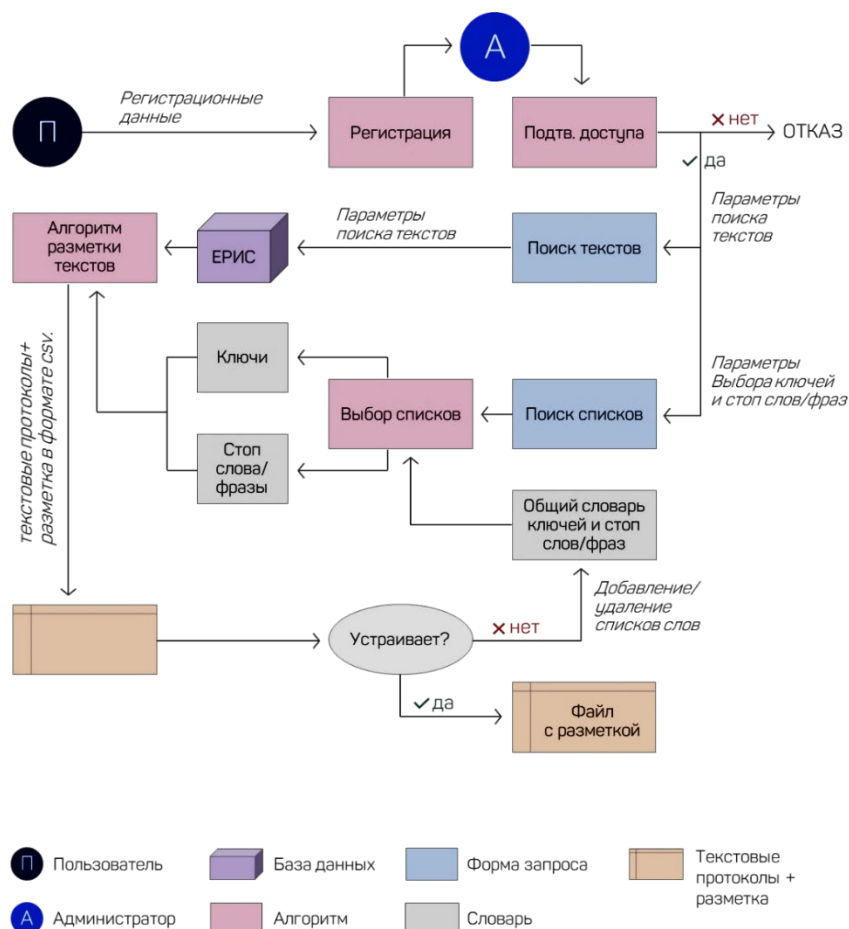


Рисунок 6 – Последовательная схема взаимодействия пользователя и NLP-инструмента

Figure 6 – Sequential diagram of interaction between the user and the NLP tool system

Данный инструмент реализует комплексную обработку медицинской документации, включающую предварительную подготовку описаний и заключений, их объединение в единый текстовый массив, формирование списка стоп-слов для посегментного анализа, создание специализированного словаря ключевых терминов для рентгенографии и флюорографии (например, «коллапс», «гидроторакс» и другие), а также дополнительные этапы обработки. Таким образом, NLP-инструмент формирует эталонные результаты (Ground truth) на основе врачебных заключений для последующего сопоставления с результатами работы СИИ. На Рисунке 6 представлены этапы работы NLP-инструмента, блоки, роли и последовательность действий (шагов).

Коэффициент эффективности C_{eff} рассчитывается по формуле:

$$C_{eff} = \frac{(T + C_{cons} + C_{AUC})}{3}, \quad (5)$$

где T – бинарная оценка времени обработки ($1 \leq 6,5$ минут, $0 > 6,5$ минут); C_{cons} – бинарный коэффициент согласия между экспертами и СИИ; C_{AUC} – бинарная оценка относительно метрики AUC-ROC.

Коэффициент согласия между экспертами и СИИ рассчитывался, в свою очередь, по следующей формуле:

$$C_{cons} = \frac{1}{2}(C_{loc} + C_{output}), \quad (6)$$

где C_{loc} – бинарный коэффициент согласия между экспертами и СИИ при локализации патологии; C_{output} – бинарный коэффициент согласия между экспертами и СИИ в выходных данных.

Результаты

Проведена оценка надежности и эффективности СИИ путем анализа реальных клинических данных, на основании разработанных методов, за десять отчетных периодов 2023 года для вида исследования РГ ОГК.

С целью сравнения численных показателей работы СИИ был проведен дисперсионный анализ (ANOVA), расчеты представлены в Таблицах 2 и 3. По результатам проведенного анализа (Таблица 2) можно сделать вывод, что для показателя «AUC» и «время работы» для различных СИИ есть значимые различия, что напрямую влияет на эффективность используемого решения.

По результатам проведенного анализа (Таблица 2) можно сделать вывод, что для всех типов дефектов есть значимые различия, что напрямую влияет на показатель надежности используемого решения.

Таблица 2 – Анализ частоты различных дефектов между СИИ за время исследования
Table 2 – Analysis of the frequency of various defects between artificial intelligence systems during the study

№ СИИ	Тип дефекта		
	Б	В	Г, Д
1	139/145542 (0,10%)	5928/145542 (4,07%)	4/640 (0,62%)
2	213/15697 (1,36%)	1176/15697 (7,49%)	47/560 (8,39%)
3	4778/241216 (1,98%)	13188/241216 (5,47%)	5/640 (0,78%)
4	1210/218610 (0,55%)	13038/218610 (5,96%)	11/800 (1,38%)
5	10/18651 (0,05%)	1388/18651 (7,44%)	10/640 (1,56%)
Суммарно	6350/639716 (0,10%)	34718/639716 (5,43%)	77/3280 (2,35%)
χ^2, p	4200, $p < 0,001$	921, $p < 0,001$	109, $p < 0,001$

Таблица 3 – Описательная статистика численных показателей работы СИИ
Table 3 – Descriptive statistics of numerical performance of artificial intelligence systems

Показатель	Параметр	СИИ					
		СИИ 1	СИИ 2	СИИ 3	СИИ 4	СИИ 5	Суммарно
AUC	N	9	15	7	8	8	47
	Mean	0,790	0,835	0,704	0,820	0,745	0,789
	SD	0,016	0,066	0,050	0,045	0,043	0,068
	95% ДИ	(0,778; 0,803)	(0,798; 0,871)	(0,658; 0,750)	(0,782; 0,857)	(0,709; 0,782)	(0,769; 0,809)
	F(ANOVA)	10,3					
	P	<0,0001					
Согласие с заключением, %	N	9	15	7	8	8	47
	Mean	74,35	84,37	75,00	80,33	79,49	79,54
	SD	9,31	8,04	11,42	3,04	13,14	9,77
	95% ДИ	(67,19; 81,51)	(79,92; 88,82)	(64,44; 85,56)	(77,79; 82,88)	(68,50; 90,48)	(76,67; 82,41)
	F(ANOVA)	2,1					
	P	0,094					
Согласие с локализацией, %	N	9	15	7	8	8	47
	Mean	74,53	82,66	76,12	79,39	80,66	79,23
	SD	9,49	8,20	11,14	3,43	11,91	9,26
	95% ДИ	(67,23; 81,82)	(78,12; 87,20)	(65,82; 86,42)	(76,53; 82,26)	(70,71; 90,62)	(76,51; 81,95)
	F(ANOVA)	1,4					
	P	0,255					
Время работы, минуты	N	9	15	7	8	8	47
	Mean	0,36	0,30	1,19	0,39	0,34	0,46
	SD	0,18	0,35	0,23	0,04	0,08	0,38
	95% ДИ	(0,22; 0,50)	(0,11; 0,49)	(0,97; 1,40)	(0,36; 0,43)	(0,28; 0,40)	(0,35; 0,58)
	F(ANOVA)	19,6					
	P	<0,0001					

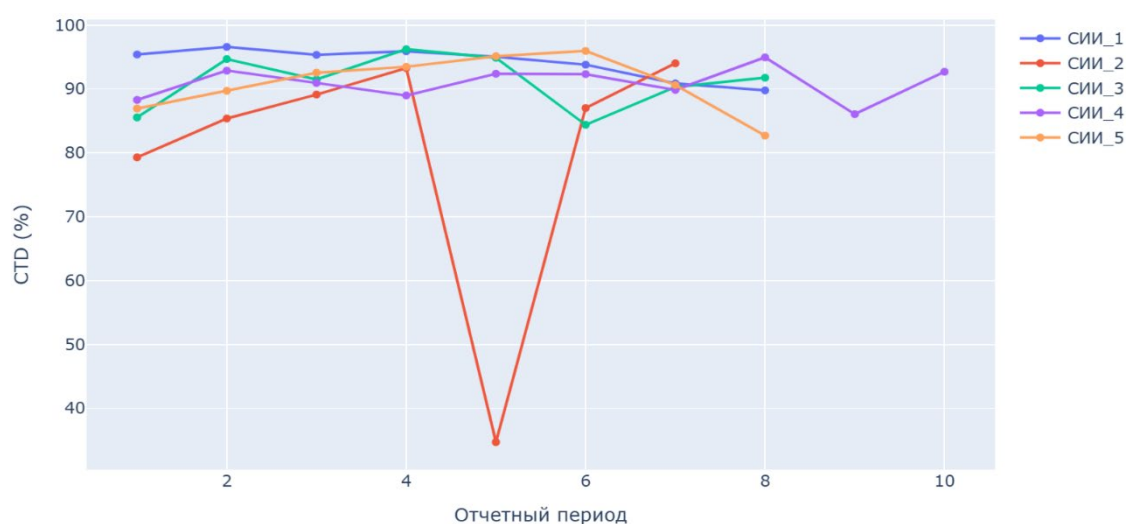


Рисунок 7 – Коэффициент надежности (Ctd) для пяти СИИ по виду исследования РГ ОГК, рассчитанный для 10 отчетных периодов

Figure 7 – Reliability coefficient (Ctd) for five artificial intelligence systems based on the type of chest X-ray examination, calculated for 10 reporting periods

На Рисунке 7 представлен график для пяти СИИ зависимости коэффициента надежности от отчетного периода (всего 10 отчетных периодов, на основании расчета и оценки технологических дефектов (Ctd)).

На основании экспериментальных данных было рассчитано среднее значение коэффициента надежности и составило 90,21 %, (95 % доверительный интервал: 67,68–99,27 %).

На Рисунке 8 представлены результаты расчетов для пяти СИИ по коэффициентам эффективности за 10 отчетных периодов (меньше, если СИИ было на доработке или смене версии) за 2023 г. Максимальным значением эффективности представляется «1», минимальным значением – «0», цветовой схемой представлены значения коэффициентов и можно проследить, как они менялись с течением времени. Знаком «v» отмечена смена версии в отчетном периоде.

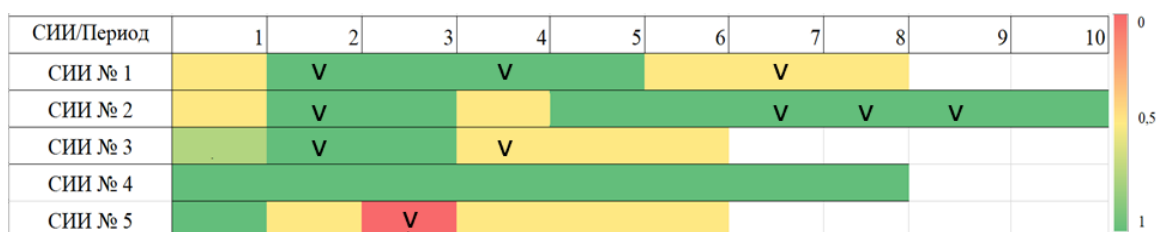


Рисунок 8 – Коэффициент эффективности для пяти СИИ для 10 отчетных периодов
Figure 8 – The effectiveness coefficient for five artificial intelligence systems for 10 reporting periods

Обсуждение

Анализ данных, представленных на Рисунках 7 и 8, демонстрирует, что при интеграции СИИ в новую операционную среду, их модификации или обновлении версий наблюдаются изменения показателей надежности и эффективности. Эти изменения могут проявляться как в положительной, так и в отрицательной динамике, либо характеризоваться относительной стабильностью. Существенно, что отсутствует прямая корреляция между версионными изменениями СИИ и динамикой исследуемых коэффициентов, что подтверждает значимость систематического мониторинга функциональности подобных систем и необходимость формирования валидированных методологических подходов к объективной оценке технологий искусственного интеллекта в медицинской практике.

В контексте необходимости совершенствования методологии мониторинга, специфичной именно для СИИ, был проведен всесторонний анализ международной научной литературы. Методологическая база разработанного подхода сформирована с учетом требований, утвержденных Коллегией Евразийской экономической комиссии, а также рекомендаций ведущих зарубежных профессиональных сообществ и регуляторных органов. Комплексная оценка характеристик программных продуктов способствует повышению информированности о текущей ситуации, предоставляя регуляторным органам и разработчикам возможность принимать обоснованные решения на основе обширной доказательной базы [8].

В период активного внедрения технологий искусственного интеллекта в клиническую практику принципиально важно обеспечить методологическую основу для непрерывной оценки их производительности, анализа потенциальных ошибок, рисков и возможного негативного воздействия. Научная литература документально подтверждает случаи снижения эффективности СИИ, обусловленные изменениями в составе анализируемых данных (включение новых клинических случаев), трансформацией моделей клинической практики [9], эволюцией терапевтических подходов,

модификацией среды применения и другими факторами [10, 11]. При этом долгосрочная перспектива использования алгоритмов искусственного интеллекта напрямую зависит от наличия регламентированных процедур, обеспечивающих регулярное обновление моделей с гарантированным сохранением их надежности и эффективности [12], а также соответствия качественным стандартам [13]³. В рамках настоящего исследования приоритетное внимание было уделено разработке и практической апробации методов оценки надежности и эффективности СИИ.

Заключение

На основании проведенного анализа можно сделать результирующий вывод, что основные усилия ИИ-разработчиков для здравоохранения должны быть смещены от достижения высоких показателей эффективности и надежности, в том числе, в сторону оценки этих метрик в формате наблюдательных исследований и оценки, технического и клинического мониторинга СИИ в условиях реальной клинической практики, что обеспечит большую стабильность в работе и надежность таких алгоритмов. Предлагаемые методы оценки надежности и эффективности СИИ в своей основе содержат стандартные подходы для ПО, при этом дополнительно обогащены показателями, отражающими специфику медицинских изделий, которыми являются СИИ в здравоохранении. Представленные методы универсальны для области лучевой диагностики и могут быть масштабированы на другие направления медицины.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Ball H.C. Improving Healthcare Cost, Quality, and Access, Through Artificial Intelligence and Machine Learning Applications. *Journal of Healthcare Management*. 2021;66(4):271–279. <https://doi.org/10.1097/JHM-D-21-00149>
2. Chen M., Decary M. Artificial Intelligence in Healthcare: An Essential Guide for Health Leaders. *Healthcare Management Forum*. 2019;33(1):10–18. <https://doi.org/10.1177/0840470419873123>
3. Vokinger K.N., Feuerriegel S., Kesselheim A.S. Continual Learning in Medical Devices: FDA's Action Plan and Beyond. *The Lancet*. 2021;3(6):E337–E338. [https://doi.org/10.1016/S2589-7500\(21\)00076-5](https://doi.org/10.1016/S2589-7500(21)00076-5)
4. Casado F.E., Lema D., Criado M.F., Iglesias R., Regueiro C.V., Barro S. Concept Drift Detection and Adaptation for Federated and Continual Learning. *Multimedia Tools and Applications*. 2022;81(3):3397–3419. <https://doi.org/10.1007/s11042-021-11219-x>
5. Pemberton H.G., Zaki L.A.M., Goodkin O., et al. Technical and Clinical Validation of Commercial Automated Volumetric MRI Tools for Dementia Diagnosis – a Systematic Review. *Neuroradiology*. 2021;63(11):1773–1789. <https://doi.org/10.1007/s00234-021-02746-3>
6. Nomura Yu., Miki S., Hayashi N., et al. Novel Platform for Development, Training, and Validation of Computer-Assisted Detection/Diagnosis Software. *International Journal of Computer Assisted Radiology and Surgery*. 2020;15(4):661–672. <https://doi.org/10.1007/s11548-020-02132-z>
7. Четвериков С.Ф., Арзамасов К.М., Андрейченко А.Е., Новик В.П., Бобровская Т.М., Владимирский А.В. Подходы к формированию выборки для контроля качества работы систем искусственного интеллекта в медико-биологических

³ Васильев Ю.А., Владимирский А.В., Омелянская О.В. и др. Оценка зрелости технологий искусственного интеллекта для здравоохранения. Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы; 2023. 28 с.

- исследованиях. *Современные технологии в медицине*. 2023;15(2):19–27. <https://doi.org/10.17691/stm2023.15.2.02>
- Chetverikov S.F., Arzamasov K.M., Andreichenko A.E., Novik V.P., Bobrovskaya T.M., Vladzimirsky A.V. Approaches to Sampling for Quality Control of Artificial Intelligence in Biomedical Research. *Modern Technologies in Medicine*. 2023;15(2):19–27. (In Russ.). <https://doi.org/10.17691/stm2023.15.2.02>
8. Harvey H.B., Gowda V. How the FDA Regulates AI. *Academic Radiology*. 2020;27(1):58–61. <https://doi.org/10.1016/j.acra.2019.09.017>
9. Sounderajah V., Ashrafian H., Golub R.M., et al. Developing a Reporting Guideline for Artificial Intelligence-Centred Diagnostic Test Accuracy Studies: the STARD-AI Protocol. *BMJ Open*. 2021;11(6). <https://doi.org/10.1136/bmjopen-2020-047709>
10. Khinvasara T., Tzenios N., Shankar A. Post-Market Surveillance of Medical Devices Using AI. *Journal of Complementary and Alternative Medical Research*. 2024;25(7):108–122. <https://doi.org/10.9734/jocamr/2024/v25i7552>
11. Lyell D., Wang Yi., Coiera E., Magrabi F. More Than Algorithms: an Analysis of Safety Events Involving ML-Enabled Medical Devices Reported to the FDA. *Journal of the American Medical Informatics Association*. 2023;30(7):1227–1236. <https://doi.org/10.1093/jamia/ocad065>
12. Feng J., Phillips R.V., Malenica I., et al. Clinical Artificial Intelligence Quality Improvement: Towards Continual Monitoring and Updating of AI Algorithms in Healthcare. *npj Digital Medicine*. 2022;5(1). <https://doi.org/10.1038/s41746-022-00611-y>
13. Васильев Ю.А., Зинченко В.В., Кудрявцев Н.Д., Михайлова А.А., Кляшторный В.Г., Владзими́рский А.В. Оценка удовлетворенности и вовлеченности врачей-рентгенологов при использовании программного обеспечения с искусственным интеллектом. *Врач и информационные технологии*. 2024;(1):70–81. https://doi.org/10.25881/18110193_2024_1_70
- Vasilev Yu.A., Zinchenko V.V., Kudryavtsev N.D., Mikhailova A.A., Klyashorny V.G., Vladzimirsky A.V. Radiologists' Satisfaction and Engagement with Artificial Intelligence Software. *Medical Doctor and IT*. 2024;(1):70–81. (In Russ.). https://doi.org/10.25881/18110193_2024_1_70

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Зинченко Виктория Валерьевна, начальник сектора клинических и технических испытаний, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Российская Федерация.
e-mail: ZinchenkoVV1@zdrav.mos.ru
ORCID: [0000-0002-2307-725X](https://orcid.org/0000-0002-2307-725X)

Victoria V. Zinchenko, Head of the Clinical and Technical Trials Branch, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation.

Владзими́рский Антон Вячеславович, доктор медицинских наук, заместитель директора по научной работе, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Российская Федерация.
e-mail: npcmr@zdrav.mos.ru
ORCID: [0000-0002-2990-7736](https://orcid.org/0000-0002-2990-7736)

Anton V. Vladzimirskyy, Doctor of Medical Sciences, Deputy Director for Scientific Work, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation.

Арзамасов Кирилл Михайлович, кандидат медицинских наук, руководитель отдела медицинской информатики, радиомики и радиогеномики, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы; доцент МИРЭА – Российский технологический университет, Москва, Российская Федерация.

e-mail: npcmr@zdrav.mos.ru

ORCID: [0000-0001-7786-0349](https://orcid.org/0000-0001-7786-0349)

Kirill M. Arzamasov, Candidate of Medical Sciences, Head of the Department of Medical Informatics, Radiomics and Radiogenomics, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department; docent MIREA – Russian Technological University, Moscow, the Russian Federation.

Статья поступила в редакцию 04.04.2025; одобрена после рецензирования 22.04.2025; принята к публикации 28.04.2025.

The article was submitted 04.04.2025; approved after reviewing 22.04.2025; accepted for publication 28.04.2025.