

DOI: <https://doi.org/10.17816/DD501759>

# Технологические дефекты программного обеспечения с искусственным интеллектом

В.В. Зинченко, К.М. Арзамасов, Е.И. Кремнева, А.В. Владзимирский, Ю.А. Васильев

Научно-практический клинический центр диагностики и телемедицины, Москва, Российская Федерация

## АННОТАЦИЯ

**Обоснование.** Технологические дефекты в работе программного обеспечения с искусственным интеллектом являются критически важными при принятии решения о практической применимости и клинической ценности программного обеспечения с искусственным интеллектом.

**Цель** — анализ и систематизация технологических дефектов, возникающих при работе программного обеспечения с искусственным интеллектом для анализа медицинских изображений.

**Материалы и методы.** В рамках эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы проводится мониторинг технологических параметров для всех участвующих решений как на этапе апробации, так и на этапе опытной эксплуатации. В статье представлена графическая информация о среднем числе технологических дефектов для профилактического направления, модальность «Маммография», за 2021 год. Этот период выбран как наиболее показательный, характеризующийся активным развитием программного обеспечения с искусственным интеллектом с позиции увеличения технической стабильности их работы. С целью оценки применимости подхода по выявлению технологических дефектов аналогичный анализ проводился для направления обнаружения внутрисерепных кровоизлияний на компьютерных томограммах головного мозга за 2022–2023 годы.

**Результаты.** В ходе исследования было проанализировано программное обеспечение с искусственным интеллектом по модальностям «Маммография» (2 алгоритма) и «Компьютерная томография головного мозга» (1). Всего для модальности «Маммография» собрано 14 выборок по 20 исследований; для модальности «Компьютерная томография» — 12 выборок по 80 исследований. Для каждого типа дефекта были построены графики, а для каждой из модальностей были построены линии тренда. Коэффициенты уравнений линий трендов указывают на тенденцию к снижению числа технологических дефектов.

**Заключение.** Проведённый анализ позволяет проследить тенденцию к снижению числа технологических дефектов, что может свидетельствовать о доработке программного обеспечения с искусственным интеллектом и повышении его качества благодаря периодическому мониторингу. Кроме того, такой результат показывает универсальность использования как для профилактических методов, так и для экстренных.

**Ключевые слова:** искусственный интеллект; программное обеспечение с искусственным интеллектом; технологический мониторинг; технологические дефекты; Московский эксперимент.

## Как цитировать:

Зинченко В.В., Арзамасов К.М., Кремнева Е.И., Владзимирский А.В., Васильев Ю.А. Технологические дефекты программного обеспечения с искусственным интеллектом // Digital Diagnostics. 2023. Т. 4, № 4 С. 593–604. DOI: <https://doi.org/10.17816/DD501759>

DOI: <https://doi.org/10.17816/DD501759>

# Technological defects in software based on artificial intelligence

Viktoriya V. Zinchenko, Kirill M. Arzamasov, Elena I. Kremneva, Anton V. Vladzemyrskyy, Yuriy A. Vasilev

Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russian Federation

## ABSTRACT

**BACKGROUND:** Technological defects in the use of artificial intelligence software are critical when deciding on the practical applicability and clinical value of artificial intelligence software.

**AIM:** To conduct an analysis and systematization of technological defects occurring when artificial intelligence software analyzes medical images.

**MATERIALS AND METHODS:** As part of the experiment on the use of innovative computer vision technologies for the analysis of medical images and further application in the Moscow healthcare system, technological parameters of all artificial intelligence software are monitored at the testing and operation stages of the trial. This article presents graphical information on the average number of technological defects in mass mammography screening in 2021. This period was chosen as the most indicative and characterized by the active development of artificial intelligence software and increased technical stability of its performance. To assess the applicability of the analysis for technological defects, a similar analysis was conducted for the direction of detection of intracranial hemorrhage on computed tomography scans of the brain for 2022–2023.

**RESULTS:** During the study, artificial intelligence software used for mammography (two algorithms) and brain computed tomography (one algorithm) were analyzed. Fourteen mammography samples were collected for technological monitoring during the identified period, each from 20 studies, and 12 brain computed tomography samples were obtained, each from 80 studies. Graphs were constructed for each type of defect, and trend lines were plotted for each modality. The coefficients of the trend line equations indicate a downward tendency in the number of technological defects.

**CONCLUSION:** This analysis allows tracing a downward trend in the number of technological defects, which may indicate a refinement of artificial intelligence software and an increase in its quality because of periodic monitoring. It also shows the versatility of use for both preventive and emergency methods.

**Keywords:** artificial intelligence; artificial intelligence software; technological monitoring; technological defects; Moscow experiment.

## To cite this article:

Zinchenko VV, Arzamasov KM, Kremneva EI, Vladzemyrskyy AV, Vasilev YuA. Technological defects in software based on artificial intelligence. *Digital Diagnostics*. 2023;4(4):593–604. DOI: <https://doi.org/10.17816/DD501759>

Received: 20.06.2023

Accepted: 21.11.2023

Published online: 06.12.2023

DOI: <https://doi.org/10.17816/DD501759>

# 人工智能软件的技术缺陷

Viktoria V. Zinchenko, Kirill M. Arzamasov, Elena I. Kremneva, Anton V. Vladzimirskyy, Yuriy A. Vasilev

Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russian Federation

## 简评

**论证。**人工智能软件性能方面的技术缺陷是确定人工智能软件实用性和临床价值的关键。

**该研究的目的是**对医学影像分析人工智能软件运行中的技术缺陷进行分析并使之系统化。

**材料和方法。**在莫斯科市进行了一项《使用创新计算机视觉技术进行医学图像分析并进一步应用于莫斯科市医疗系统的实验》。在实验框架内，对所有参与解决方案的技术参数进行监测。监测是在批准阶段和试运行阶段进行的。本文以图表形式介绍2021年“乳房摄影术”预防方向的平均技术缺陷数量。这一时期被选为最有意义的时期。这一时期的特点是从提高操作技术稳定性的角度出发，积极开发人工智能软件。为了评估该方法在发现技术缺陷方面的适用性，我们对2022–2023年脑部CT扫描颅内出血的检测方向进行了类似的分析。

**结果。**本研究分析了“乳房摄影术”（2种算法）和“脑计算机断层扫描”（1种）模式的人工智能软件。在“乳房X射线照相术”模式中，共收集了14个样本，共有20项研究。在“脑计算机断层扫描”模式中，共收集了12个样本，共有80项研究。我们对每种缺陷类型都绘制了图表，对每种模式绘制了趋势线。趋势线公式的系数表明了，技术缺陷的数量呈下降趋势。

**结论。**通过分析，我们发现了减少技术缺陷数量的趋势。这可能表明人工智能软件的完善，以及通过定期监测，软件质量的提升。此外，这一结果还显示使用预防和应急方法的通用性。

**关键词：**人工智能；人工智能软件；技术监测；技术缺陷；莫斯科实验。

## 引用本文：

Zinchenko VV, Arzamasov KM, Kremneva EI, Vladzimirskyy AV, Vasilev YuA. 人工智能软件的技术缺陷. *Digital Diagnostics*. 2023;4(4):593–604. DOI: <https://doi.org/10.17816/DD501759>

收到: 20.06.2023

接受: 21.11.2023

发布日期: 06.12.2023

## ОБОСНОВАНИЕ

Программное обеспечение с технологиями искусственного интеллекта (ПО с ИИ) может помочь в рутинных и сложных задачах медицинскому персоналу, а также повысить уровень, доступность и скорость оказываемой пациентам медицинской помощи [1–3]. Это стало возможным в большей степени благодаря «Эксперименту по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы» (далее — Эксперимент), а также преемственности зарубежного и отечественного опыта по работе с искусственным интеллектом в здравоохранении [4–7]. Цель Эксперимента — научное исследование возможности использования в системе здравоохранения города Москвы методов поддержки принятия врачебных решений на основе результатов анализа данных с применением передовых инновационных технологий. Были разработаны требования к результатам работы ПО с ИИ по 21 направлению лучевой диагностики. На данный момент врачам доступны результаты более 50 решений на основе искусственного интеллекта. По состоянию на конец сентября 2023 обработано более 10 млн исследований.

Применение новых технологий в здравоохранении требует обязательного соблюдения правил безопасности, поэтому этапы разработки, ввода в эксплуатацию и применения ПО с ИИ должны в обязательном порядке контролироваться [8]. ПО с ИИ требует особого контроля в процессе эксплуатации ещё и по той причине, что в нём возможно смещение результатов при его использовании на отличной от обучающей популяции [9, 10].

Для контроля качества обработки исследований ПО с ИИ в рамках Эксперимента используют ряд тестирований. Самотестирование — начальный этап, предназначен для понимания технической совместимости ПО с ИИ и подаваемых на обработку исследований (входных данных). Следующий этап — функциональное тестирование, которое определяет наличие заявленных функций ПО с ИИ, в том числе и его работоспособность. Оценка ПО с ИИ проводится как с технической, так и с клинической точек зрения, то есть техническими и медицинскими экспертами. Калибровочное тестирование — этап определения метрик работы ПО с ИИ, где основной показатель — площадь под ROC-кривой.

В случае успешного прохождения всех тестирований, ПО с ИИ допускается к работе, при этом по результатам работы для алгоритмов проводят технологический и клинический мониторинги. Техническое тестирование (мониторинг технологических параметров), согласно международным исследованиям, — неотъемлемая часть проверки продукта, которую проводят в рамках комплексного тестирования для возможности использования в условиях реальной клинической практики [11]. Именно поэтому в рамках данной работы мы сосредоточились на мониторинге технологических дефектов.

## ЦЕЛЬ

Изучение технологических дефектов ПО с ИИ по результатам Эксперимента, их анализ, статистика и влияние на безопасность и качество ПО с ИИ в условиях работы в клинической практике.

## МАТЕРИАЛЫ И МЕТОДЫ

### Условия проведения

Мониторинг выполнялся для всех исследований, проанализированных ПО с ИИ за отчётный период для модальности «Маммография» в 2021 г., по категориям дефектов, указанных в табл. 1 (левая колонка, согласно Приказу Департамента здравоохранения города Москвы № 51 от 26.01.2021 г.) [12]:

- **Группа «а»** — время, затрачиваемое на анализ одного исследования, превышает 6,5 минут. Ограничение по времени было получено эмпирическим путём как среднее время описания исследования ПО с ИИ для того, чтобы врач-рентгенолог смог воспользоваться его результатами.
- **Группа «б»** — отсутствие результатов проанализированных исследований.
- **Группа «в»** — изображения, содержащиеся в результатах работы ПО с ИИ, не соответствуют изображениям нативного (исходного) исследования (искажены). В редких случаях изменение метаданных может приводить к изменению настроек при просмотре исследований, что существенно затрудняет визуализацию оригинального изображения.
- **Группа «г»** — некорректная работа заявленного функционала ПО с ИИ, затрудняющая работу врача или делающая её выполнение невозможным с надлежащим качеством, в том числе — обрезка изображений, изменение яркости/контрастности, отсутствие описания результата, отсутствие разметки при наличии патологии.
- **Группа «д»** — иные нарушения целостности и содержимого файлов с результатами исследований, обуславливающих ограничение их диагностической интерпретации, в том числе — разметка за пределами целевого органа, анализ ПО с ИИ некорректной анатомической области.
- **Группа «е»** — изменение оригинальной серии исследования. В 2022 г. была произведена реструктуризация дефектов, что учтено при обработке данных мониторингов для модальности «КТ» (см. табл. 1, правая колонка).

### Продолжительность исследования

Мониторинг проводился ежемесячно до момента окончания участия ПО с ИИ в рамках работы Эксперимента. Отчётный период мониторинга равен календарному месяцу.

**Таблица 1.** Соотношение критериев технологических дефектов в Приказах Департамента здравоохранения города Москвы

Технологические дефекты согласно Приказу Департамента здравоохранения города Москвы № 51 от 26.01.2021 г. (актуальные для данных маммографии, представленных в статье)	Технологические дефекты согласно Приказу Департамента здравоохранения города Москвы № 160 от 03.11.2022 г. (реструктурированные) (актуальные для данных КТ головного мозга, представленных в статье)
группа «а» — время, затрачиваемое на анализ одного исследования, превышает 6,5 минут	группа «а» — время, затрачиваемое на анализ одного исследования, превышает 6,5 минут
группа «б» — отсутствие результатов проанализированных исследований	группа «б» — отсутствие результатов проанализированных исследований
—	группа «в» — некорректная работа заявленного функционала ПО с ИИ, затрудняющая работу врача-рентгенолога или делающая её выполнение невозможным с надлежащим качеством
г2 — отсутствие дополнительной серии	в1 — отсутствие дополнительной серии
г3 — отсутствие DICOM SR	в2 — отсутствие DICOM SR
г4 — наличие двух и более DICOM SR	в3 — наличие двух и более DICOM SR
г5 — отсутствие названия ПО с ИИ	в4 — отсутствие названия ПО с ИИ
г6 — отсутствие сведений о версии ПО с ИИ	в5 — отсутствие сведений о версии ПО с ИИ
—	группа «г» — дефекты, связанные с отображением области изображений
в1 — изображения обрезаны	г1 — изображения в дополнительной серии обрезаны
в2 — изменена яркость/контрастность	г2 — яркость/контрастность дополнительной серии не соответствует оригинальному изображению
в3 — проанализированы не все необходимые изображения	г3 — проанализированы не все необходимые изображения
г1 — полное отсутствие результатов работы ПО с ИИ	исключено
г7 — отсутствие предупреждающей надписи: «Только для использования в исследовательских/научных целях»	г4 — отсутствие предупреждающей надписи: «Только для использования в исследовательских/научных целях»
г8 — отсутствие разметки при наличии патологии	«е» — дефекты, связанные с клинической работой
Д1 — противоречие информации DICOM SR и дополнительной серии	исключено
Е — изменение оригинальной серии исследования	г5 — изменение оригинальной серии исследования
—	группа «д» — иные нарушения целостности и содержимого файлов с результатами исследований, обуславливающие ограничение их диагностической интерпретации в том числе
Д2 — разметка за пределами целевого органа	д1 — разметка за пределами целевого органа
Д3 — проанализирована некорректная анатомическая область, проекция или серия	д2 — проанализирована некорректная анатомическая область, проекция или серия

Примечание. SR — structure report.

Промежуточный отчёт по мониторингу дефектов групп «а» и «б» формировался по данным на 10-е и 20-е число каждого месяца и направлялся производителю ПО с ИИ.

Для выборок маммографических исследований в статье указана информация о дефектах с марта по декабрь 2021 года, для выборки КТ-исследований головного мозга — с мая 2022 по май 2023 года. Для различных ПО с ИИ периодичность проводимого мониторинга различна ввиду варьирования момента захода в Эксперимент и времени доработки после получения обратной связи.

Технологический мониторинг выполнялся экспертной группой, в которую входили технические специалисты

и врачи-рентгенологи, прошедшие дополнительное обучение по проведению мониторингов и инструктаж по работе с конкретными ПО с ИИ. Для отчётности по проведённому мониторингу также была разработана и применялась унифицированная форма внутреннего отчёта и инструкция по проведению технологического мониторинга.

### Статистический анализ

Для тестирования в ходе проведения технологического мониторинга использовался псевдослучайно отобранный набор данных (выборка исследований) с учётом

следующей пропорции: 25% исследований, в которых ПО с ИИ не обнаружило патологию (группа «без патологии»), и 75% исследований с выявленной патологией (группа «с патологией»). Отобранные исследования с результатами от ПО с ИИ изучались на наличие технологических дефектов. Исследование определялось в группу «с патологией» в случае превышения оптимального порогового значения, установленного при проведении тестирования; в противном случае классифицировалось как «без патологии» [13, 14].

В 2021 году в Эксперименте объём псевдослучайной выборки составлял 20 исследований ежемесячно. При этом необходимо отметить, что это был ещё Пилотный этап проекта, и уровень мощности, согласно номограмме, составлял 42,5% при уровне статистической значимости 0,05. Стандартные различия между элементами выборки — 0,79 [15]. В полноценном проекте после 2021 года, используя риск-анализ, объём выборки составил 80 исследований (обоснование в статье С.Ф. Четверикова и соавт. [13]). Именно такой объём — 80 исследований — составлял ежемесячную выборку для КТ-исследований головного мозга.

## РЕЗУЛЬТАТЫ

Для проведения технологического мониторинга для модальности «Маммография» всего было использовано 14 выборок по 20 исследований. Сформированные псевдослучайные выборки ежемесячно во временном интервале с марта по декабрь 2021 года направлялись всем работающим (не находящимся на доработке) ПО с ИИ для тестирования.

С целью оценки применимости метода выявления технологических дефектов аналогичный анализ проводился на сформированных псевдослучайных выборках для модальности «КТ головного мозга» на предмет обнаружения внутричерепных кровоизлияний. С мая 2022 по май

2023 года в ПО с ИИ для тестирования ежемесячно направлялись 80 исследований (всего 12 выборок по 80 исследований).

Для построения графика динамики технологических дефектов за период с марта по декабрь 2021 года (для модальности «Маммография») или с мая 2022 по май 2023 года (для модальности «КТ») была использована общая статистика по всем результатам технологического мониторинга ПО с ИИ. Количество технологических дефектов считалось в процентах от общего числа исследований в наборе данных.

На рис. 1 представлена динамика среднего числа технологических дефектов для модальности «Маммография» с марта 2021 года по декабрь 2021 года, где ось ординат — наличие дефектов (представлено в процентном соотношении от общего числа исследований выборки), а ось абсцисс — отчётный период в месяцах. На рис. 2 представлена аналогичная информация для модальности «КТ головного мозга» (период с мая 2022 по май 2023 года).

Для модальности «Маммография» дефекты представлены согласно левой колонке в табл. 1. Из рис. 1 видно, что в начале исследуемого периода мониторинга преобладали дефекты групп «в» (дефекты, связанные с отображением области изображений), «г» (некорректная работа заявленного функционала ПО с ИИ) и «б» (отсутствие результатов проанализированных исследований). К концу периода исследования остались только дефекты группы «в», хотя и их доля также значительно снизилась в процентном соотношении.

Для модальности «КТ» дефекты представлены согласно правой колонке в табл. 1. Из рис. 2 для модальности «КТ головного мозга» на предмет обнаружения внутричерепного кровоизлияния видно, что процент соотношения дефектов к выборке для всех дефектов, кроме группы «б» (отсутствие результатов проанализированных

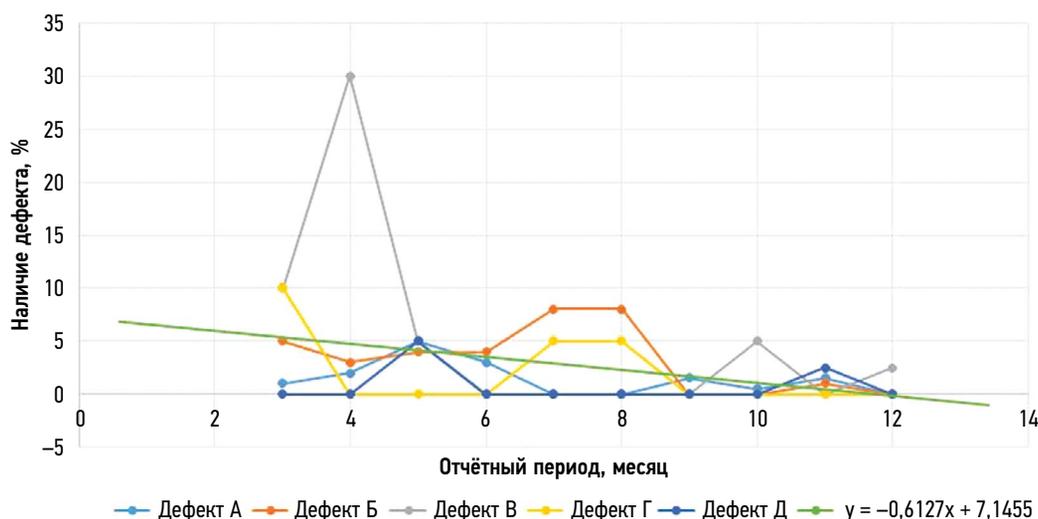
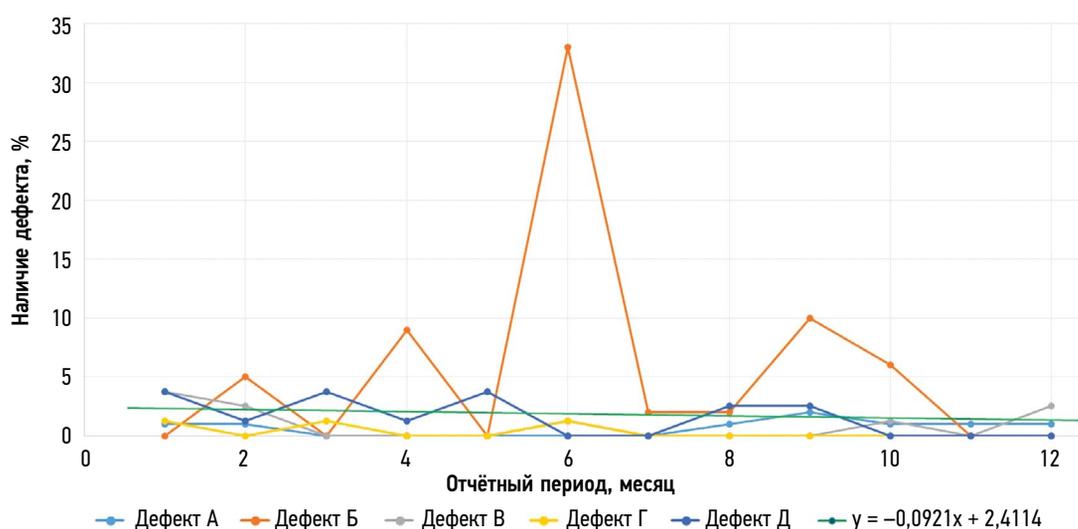


Рис. 1. Динамика выявления среднего числа каждого технологического дефекта для программного обеспечения с искусственным интеллектом модальностью «Маммография». Дефекты распределены по группам согласно Приказу Департамента здравоохранения города Москвы № 51 от 26.01.2021 г.



**Рис. 2.** Динамика выявления числа каждого технологического дефекта для программного обеспечения с искусственным интеллектом модальностью «Компьютерная томография головного мозга» (наличие или отсутствие внутричерепного кровоизлияния). Дефекты распределены по группам согласно Приказу Департамента здравоохранения города Москвы № 160 от 03.11.2022 г.

исследований), был ниже, чем для других проанализированных модальностей. При этом отмечалось снижение (в процентном соотношении) количества дефектов групп «Г» (дефекты, связанные с отображением области изображений) и «Д» (нарушения целостности и содержимого файлов), тогда как дефекты группы «Б» показали большой разброс от месяца к месяцу.

С целью количественной оценки данной тенденции были добавлены соответствующие линии тренда. Это линейные функции вида  $k \cdot x + b$ , где коэффициент  $k$  задаёт наклон кривой аппроксимации, то есть указывает на тенденцию к росту или снижению числа дефектов, а коэффициент  $b$  соответствует количеству дефектов в начале мониторинга. Аппроксимация проводилась для всех ПО с ИИ по отдельным модальностям, при этом аппроксимировался сразу весь массив данных (см. рис. 1–2). Зная коэффициент  $k$ , можно прогнозировать динамику в ликвидации технологических дефектов для каждого ПО с ИИ в отдельности или для целого направления.

На рис. 3–6 представлены примеры технологических дефектов для ПО с ИИ.

## ОБСУЖДЕНИЕ

На основании полученных и проанализированных результатов исследования, модальность «Маммография» показывает отличную тенденцию по снижению числа технологических дефектов (см. рис. 1 — линия тренда). ПО с ИИ с модальностью «КТ головного мозга» имеет более ровную тенденцию к снижению числа технологических дефектов (см. рис. 2 — линия тренда), несмотря на значения дефектов группы «Б». Это, в свою очередь, можно объяснить тем, что флуктуации некоторых технологических дефектов связаны с автоматическим выявлением, а также быстрой обратной связью и оперативной

доработкой ПО с ИИ производителем (сменой версии или исправлением ошибок).

Важно отметить, что смена версий ПО с ИИ для модальности «Маммография» пришлась на сентябрь–октябрь 2021 года, и в это же время снизилось в среднем количество дефектов групп «Б» и «Г» (см. рис. 1). Это может указывать на успешную доработку ПО с ИИ, что, в свою очередь, может свидетельствовать об эффективности применения представленной методологии технологического мониторинга.

Необходимо отметить, что выявление технологических дефектов в рамках проведения технологического мониторинга представляется важной частью комплексного тестирования с целью более безопасной, качественной и эффективной работы ПО с ИИ не только в области лучевой диагностики, но и в целом в медицине. Результаты проведённого анализа показывают, что с уменьшением количества дефектов качество ПО с ИИ возрастает. Таким образом, ПО с ИИ получают больше доверия от пользователей, работая с минимальным количеством дефектов и помогая врачу [16, 17].

## Реструктурирование технологических дефектов

На основании результатов мониторинга технологических дефектов и их анализа, приведённых в данной статье, в 2022 году были реструктурированы группы технологических дефектов. Мониторинг дефектов ПО с ИИ для модальности «КТ головного мозга» на предмет наличия или отсутствия внутричерепного кровоизлияния был проведён в соответствии с обновлённой классификацией групп (см. табл. 1, правая колонка). Дефекты групп «А» и «Б» проходили проверку в автоматизированном режиме, а дефекты групп «В», «Г» и «Д» требовали ручной проверки от экспертов. Обновлённый перечень

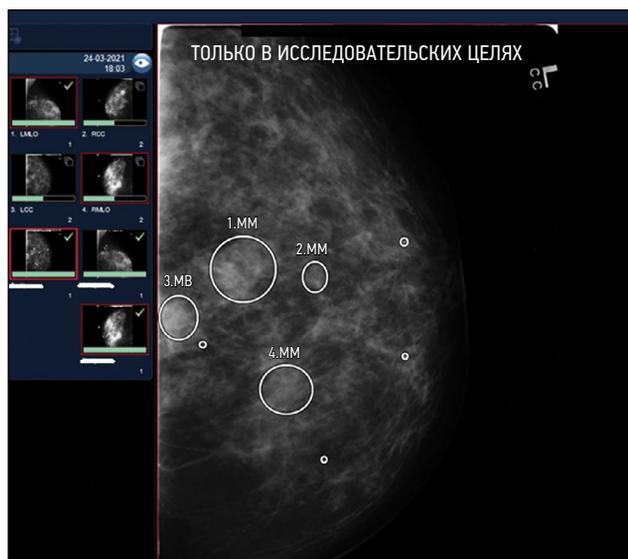


Рис. 3. Дефект — проанализированы не все необходимые изображения. Модальность "Маммография".

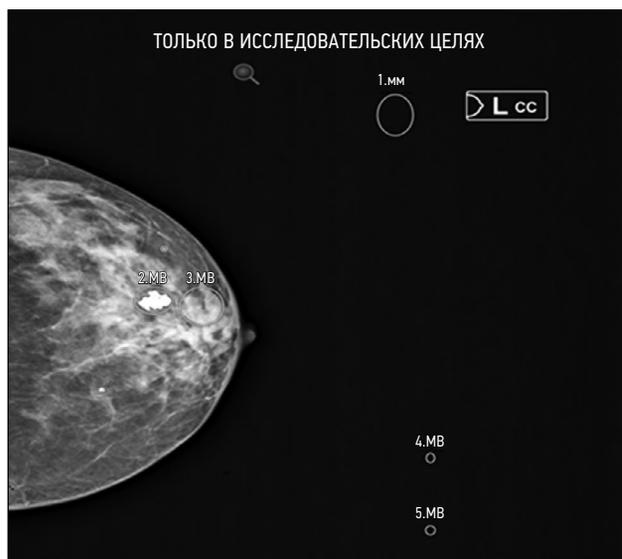


Рис. 4. Дефект — разметка за пределами целевого органа. Модальность "Маммография".

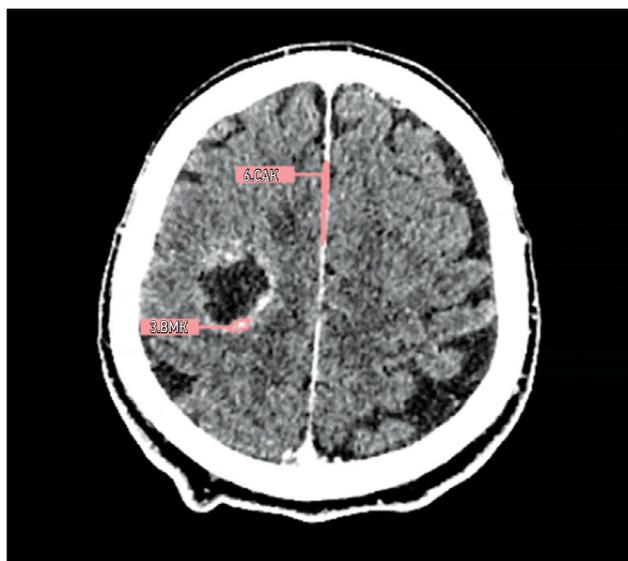


Рис. 5. Дефект — проанализирована некорректная серия (постконтрастная КТ вместо нативной). Модальность "Компьютерная томография".

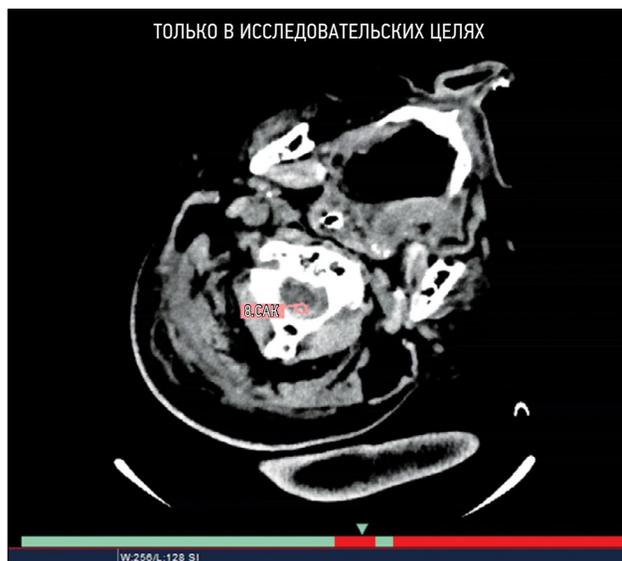


Рис. 6. Дефект — разметка за пределами целевого органа, постконтрастная компьютерная томография вместо нативной. Модальность "Компьютерная томография".

технологических дефектов представлен в Приказе Департамента здравоохранения города Москвы № 160 от 03.11.2022 г. и актуален по настоящее время [18]. Обоснование увеличения числа исследований в выборке представлено в статье С.Ф. Четверикова и соавт. [13]. Подобная реструктуризация технологических дефектов позволила оптимизировать работу экспертов, анализирующих результаты мониторингов ПО с ИИ.

Кроме того, по итогу проведения технологического мониторинга ПО с ИИ в условиях Эксперимента, технологические дефекты согласно Приказу от 2021 г. для модальности «Маммография» можно условно разделить на три группы, касательно безопасности ПО с ИИ как медицинского изделия:

- **Влияют на безопасность пациента и на работу врача** — отсутствие реализации заявленного функционала от производителя; замечания, оказывающие влияние на врача-рентгенолога или затрудняющие его деятельность; необратимое искажение исходных данных исследования. К этой группе относятся, например, дефекты «г» («г2», «г3», «г4»), а также дефект группы «е» (изменение оригинальной серии исследования, что напрямую влияет на безопасность использования ПО с ИИ). Отдельно необходимо рассмотреть дефект «г7» (отсутствие предупредительной надписи: «Только для использования в исследовательских/научных целях»). Этот дефект может иметь место только

в рамках научного исследования, но ни в коем случае не в рамках использования ПО с ИИ как медицинского изделия.

- **Не влияют на безопасность пациента, но влияют на работу врача** — недоработки функционала, которые не соответствуют общепринятым стандартам представления результатов интерпретации исследования. К этой группе можно отнести дефекты групп «д» и «в» («в1», «в2», «в3»).
- **Не влияют на безопасность пациента, не влияют на работу врача** — несущественные недоработки, которые необходимо устранить для более удобной, интуитивно понятной и оперативной работы врача. К данной группе можно отнести дефекты «г5», «г6» и «г8».

Для модальности «КТ» ввиду реструктуризации дефектов (табл. 1) разделение по трём подгруппам касательно безопасности, начиная с ноября 2023 г. и до данного периода, представлялось согласно Приказу 2021 г.:

- **Влияют на безопасность пациента и на работу врача:** дефекты «в» («в1», «в2», «в3»), а также дефект «г4», «г5».
- **Не влияют на безопасность пациента, но влияют на работу врача:** дефекты групп «д» и «г» («г1», «г2», «г3»).
- **Не влияют на безопасность пациента, не влияют на работу врача:** дефекты «в4» и «в5».

Графически информация об изменении количества дефектов по группам и по месяцам представлена на рис. 7–8 для обеих модальностей.

Для модальности «Маммография» (см. рис. 7) дефекты, влияющие на безопасность пациента и на работу врача, в результате доработки ПО с ИИ не выявляются после июня. Дефекты, влияющие на работу врача, но не влияющие на безопасность пациента, также имеют тенденцию к снижению к моменту окончания периода исследования.

Для модальности «КТ головного мозга» (см. рис. 8) наиболее часто встречающиеся дефекты (влияющие на работу врача, но не влияющие на безопасность пациента) не имеют явной тенденции к снижению.

Представленная в данной статье методология мониторинга технологических дефектов позволяет осуществлять контроль технической стабильности алгоритмов, что имеет важное практическое значение при оценке ПО с ИИ, его безопасности. Применяемая методология контроля работы ПО с ИИ на потоке позволила выявить технологические дефекты и доработать решения, что в конечном итоге привело к повышению технологической стабильности ПО с ИИ, что мы видим на примере данных анализа КТ головного мозга. Таким образом, разработанная методология оказалась эффективным и универсальным инструментом повышения технической стабильности ПО с ИИ.

## ЗАКЛЮЧЕНИЕ

В настоящей работе представлен перечень основных технологических дефектов, возникающих при внедрении ПО с ИИ, а также методология мониторинга технологических дефектов, основанная на регулярном выборочном

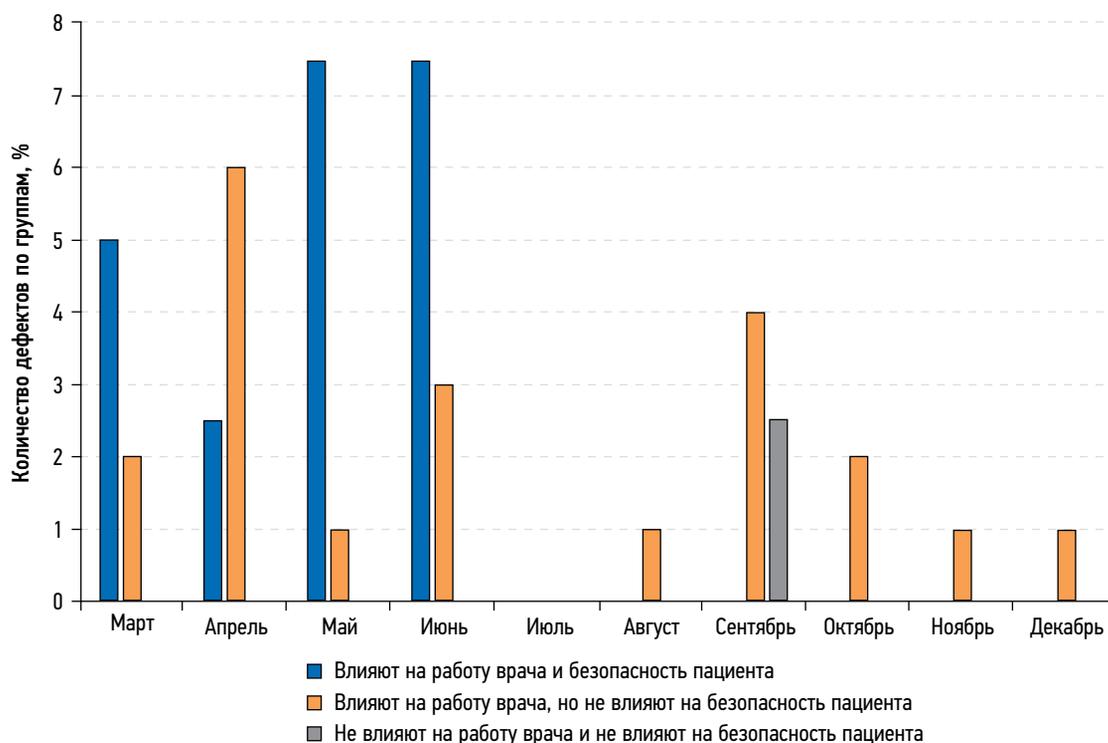


Рис. 7. Количество дефектов в группе в динамике для модальности «Маммография».

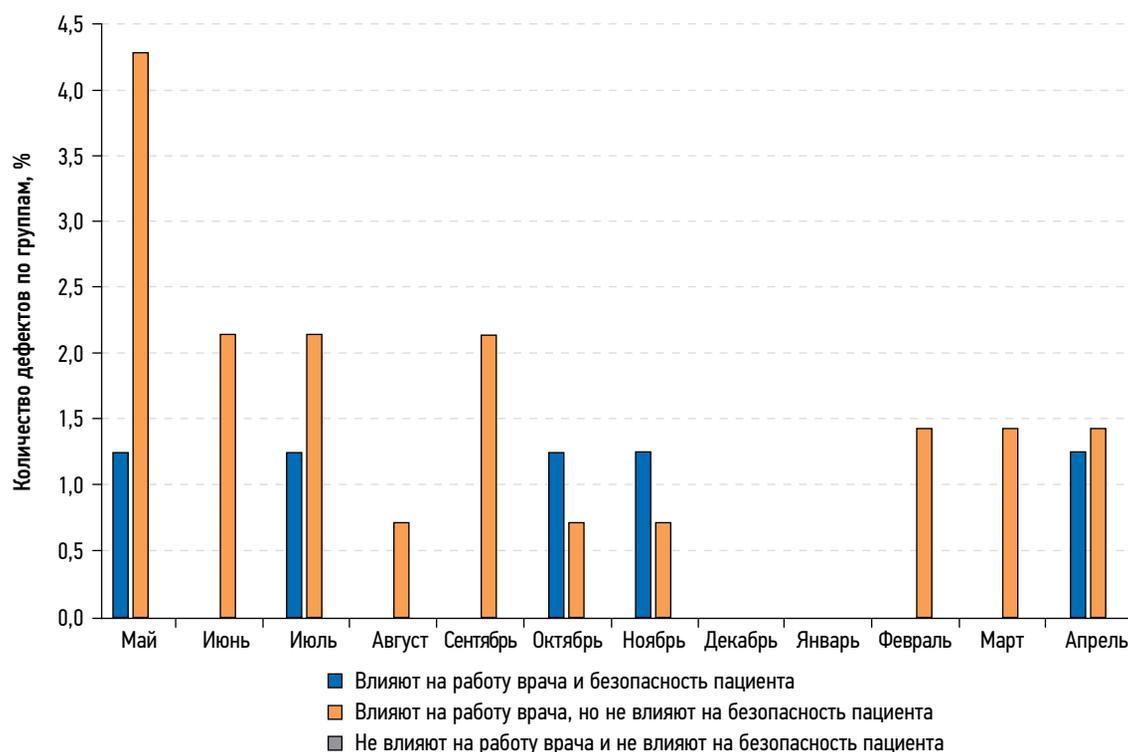


Рис. 8. Количество дефектов в группе в динамике для модальности «Компьютерная томография».

контрольном тестировании, позволяющая повышать техническую стабильность ПО с ИИ. Разработанная методология тестирования ПО с ИИ при выявлении технологических дефектов представляется частью мониторинга безопасности, качества и эффективности для тестирования ПО с ИИ в условиях работы в реальной клинической практике.

## ДОПОЛНИТЕЛЬНО

**Источник финансирования.** Анализ и оценка технологических дефектов КТ-выборки на предмет наличия или отсутствия внутричерепного кровоизлияния подготовлено за счёт гранта Российского научного фонда № 22-25-20231, <https://rscf.ru/project/22-25-20231/>.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Вклад авторов.** Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение работы и подготовку статьи, прочли и одобрили финальную версию перед публикацией).

Наибольший вклад распределён следующим образом: В.В. Зинченко — структурирование и анализ полученных результатов, написание рукописи статьи; К.М. Арзамасов — получение данных технологического мониторинга, анализ полученных результатов, корректировка рукописи статьи;

Е.И. Кремнева — структурирование и анализ полученных результатов (компьютерная томография), написание текста статьи; А.В. Владимирский — рецензирование рукописи статьи, формирование гипотезы исследования; Ю.А. Васильев — формирование гипотезы исследования, общее управление исследованием.

## ADDITIONAL INFORMATION

**Funding source.** Analysis of technological defects in computed tomography dataset with or without intracranial bleeding was funded by Russian Science Foundation Grant № 22-25-20231, <https://rscf.ru/project/22-25-20231/>.

**Competing interests.** The authors declare that they have no competing interests.

**Authors' contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. The contribution is distributed as follows: V.V. Zinchenko — structuring and analysis of the results obtained (mammography modality), writing the manuscript of the article; K.M. Arzamasov — obtaining technological monitoring data, analyzing the results obtained, correcting the manuscript of the article; E.I. Kremneva — structuring and analysis of the results obtained (computed tomography modality), writing the manuscript of the article; A.V. Vladzmyrskyy — review of the manuscript of the article, formation of the research hypothesis; Yu.A. Vasilev — formation of the research hypothesis, general management of the research.

## СПИСОК ЛИТЕРАТУРЫ

1. Владимирский А.В., Васильев Ю.А., Арзамасов К.М. и др. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента / под ред. Ю.А. Васильева, А.В. Владимирского. Издательские решения, 2022.
2. Ranschaert E.R., Morozov S., Algra P.R., editors. *Artificial Intelligence in Medical Imaging*. Berlin : Springer, 2019. doi: 10.1007/978-3-319-94878-2
3. Гусев А.В., Добридюк С.Л. Искусственный интеллект в медицине и здравоохранении // Информационное общество. 2017. № 4-5. С. 78–93.
4. Шутов Д.В., Шарова Д.Е., Абуладзе Л.Р., Дроздов Д.В. Системы искусственного интеллекта в клинической физиологии: как сделать их обучение эффективным? // Digital Diagnostics. 2023. Т. 4, № 1. С. 81–88. doi: 10.17816/DD123559
5. Мелдо А.А., Уткин Л.В., Трофимова Т.Н. Искусственный интеллект в медицине: современное состояние и основные направления развития интеллектуальной диагностики // Лучевая диагностика и терапия. 2020. Т. 11, № 1. С. 9–17. doi: 10.22328/2079-5343-2020-11-1-9-17
6. Recht M.P., Dewey M., Dreyer K., et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations // *European radiology*. 2020. Vol. 30, N 6. P. 3576–3584. doi: 10.1007/s00330-020-06672-5
7. Larson D.B., Harvey H., Rubin D.L., et al. Regulatory Frameworks for Development and Evaluation of Artificial Intelligence-Based Diagnostic Imaging Algorithms: Summary and Recommendations // *Journal of the American College of Radiology*. 2021. Vol. 18, N 3 Pt A. P. 413–424. doi: 10.1016/j.jacr.2020.09.060
8. Zinchenko V., Chetverikov S., Ahmad E., et al. Changes in software as a medical device based on artificial intelligence technologies // *International Journal of Computer Assisted Radiology and Surgery*. 2022. Vol. 17. P. 1969–1977. doi: 10.1007/s11548-022-02669-1
9. Nomura Y., 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. Vol. 15, N 4. P. 661–672. doi: 10.1007/s11548-020-02132-z
10. Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы ФГБУ «ВНИИ-ИМТ» Росздравнадзора. Москва, 2021.
11. 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. Vol. 63. P. 1773–1789. doi: 10.1007/s00234-021-02746-3
12. Приказ Департамента здравоохранения города Москвы № 51 от 26.01.2021 «Об утверждении порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы в 2021 году»
13. Четвериков С.Ф., Арзамасов К.М., Андрейченко А.Е., и др. Подходы к формированию выборки для контроля качества работы систем искусственного интеллекта в медико-биологических исследованиях // *Современные технологии в медицине*. 2023. Т. 15, № 2. С. 19. doi: 10.17691/stm2023.15.2.02
14. Зинченко В.В., Арзамасов К.М., Четвериков С.Ф., и др. Методология проведения пострегистрационного клинического мониторинга для программного обеспечения с применением технологий искусственного интеллекта // *Современные технологии в медицине*. 2022. Т. 14, № 5. С. 15–25. doi: 10.17691/stm2022.14.5.02
15. Altman D.G. Statistics and ethics in medical research: III How large a sample? // *British medical journal*. 1980. Vol. 281, N 6251. P. 1336. doi: 10.1136/bmj.281.6251.1336
16. Тыров И.А., Васильев Ю.А., Арзамасов К.М., и др. Оценка зрелости технологий искусственного интеллекта для здравоохранения: методология и ее применение на материалах Московского Эксперимента по компьютерному зрению в лучевой диагностике // *Врач и информационные технологии*. 2022. № 4. С. 76–92. doi: 10.25881/18110193\_2022\_4\_76
17. Владимирский А.В., Гусев А.В., Шарова Д.Е., и др. Методика оценки уровня зрелости информационной системы для здравоохранения // *Врач и информационные технологии*. 2022. № 3. С. 68–84. doi: 10.25881/18110193\_2022\_3\_68
18. Приказ Департамента здравоохранения города Москвы от 03.11.2022 № 160 «Об утверждении порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы в 2022 году».

## REFERENCES

1. Vladzimirskii AV, Vasil'ev YuA, Arzamasov KM, et al. *Computer Vision in Radiologic Diagnostics: the First Stage of Moscow experiment*. Vasil'ev YuA, Vladzimirskii AV, editors. Publishing solutions; 2022. (In Russ).
2. Ranschaert ER, Morozov S, Algra PR, editors. *Artificial Intelligence in Medical Imaging*. Berlin: Springer; 2019. doi: 10.1007/978-3-319-94878-2
3. Gusev AV, Dobridnyuk SL. Artificial intelligence in medicine and healthcare. *Information Society Journal*. 2017;(4-5):78–93. (In Russ).
4. Shutov DV, Sharova DE, Abuladze LR, Drozdov DV. Artificial intelligence in clinical physiology: How to improve learning agility. *Digital Diagnostics*. 2023;4(1):81–88. doi: 10.17816/DD123559
5. Meldo AA, Utkin LV, Trofimova TN. Artificial intelligence in medicine: current state and main directions of development of the intellectual diagnostics. *Diagnostic radiology and radiotherapy*. 2020;11(1):9–17. doi: 10.22328/2079-5343-2020-11-1-9-17
6. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *European radiology*. 2020;30(6):3576–3584. doi: 10.1007/s00330-020-06672-5
7. Larson DB, Harvey H, Rubin DL, et al. Regulatory Frameworks for Development and Evaluation of Artificial Intelligence-Based Diagnostic Imaging Algorithms: Summary and Recommendations. *Journal of the American College of Radiology*. 2021;18(3 Pt A):413–424. doi: 10.1016/j.jacr.2020.09.060
8. Zinchenko V, Chetverikov S, Ahmad E, et al. Changes in software as a medical device based on artificial intelligence technologies. *International Journal of Computer*

*Assisted Radiology and Surgery*. 2022;17:1969–1977. doi: 10.1007/s11548-022-02669-1

9. Nomura Y, 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. doi: 10.1007/s11548-020-02132-z

10. Methodological recommendations on the procedure for expert examination of quality, efficiency and safety of medical devices (in terms of software) for state registration under the national system FGBU «VNIИИМТ» Roszdravnadzor. Moscow; 2021. (In Russ).

11. Pemberton HG, Zaki LAM, Goodkin O, et al. Technical and clinical validation of commercial automated volumetric MRI tools for dementia diagnosis — a systematic review. *Neuroradiology*. 2021;63:1773–1789. doi: 10.1007/s00234-021-02746-3

12. Order of the Moscow City Health Department No. 51 dated 26.01.2021 «On approval of the procedure and conditions for conducting an experiment on the use of innovative technologies in the field of computer vision to analyze medical images and further application in the health care system of the city of Moscow in 2021». (In Russ).

13. Chetverikov SF, Arzamasov KM, Andreichenko AE, et al. Approaches to Sampling for Quality Control of Artificial Intelligence in Biomedical Research. *Modern Technologies in Medicine*. 2023;15(2):19. doi: 10.17691/stm2023.15.2.02

14. Zinchenko VV, Arzamasov KM, Chetverikov SF, et al. Methodology for Conducting Post-Marketing Surveillance of Software as a Medical Device Based on Artificial Intelligence Technologies. *Modern Technologies in Medicine*. 2022;14(5):15–25. doi: 10.17691/stm2022.14.5.02

15. Altman DG. Statistics and ethics in medical research: III How large a sample? *British medical journal*. 1980;281(6251):1336. doi: 10.1136/bmj.281.6251.1336

16. Tyrov IA, Vasilyev YuA, Arzamasov KM, et al. Assessment of the maturity of artificial intelligence technologies for healthcare: methodology and its application based on the use of innovative computer vision technologies for medical image analysis and subsequent applicability in the healthcare system of Moscow. *Medical Doctor and IT*. 2022;(4):76–92. doi: 10.25881/18110193\_2022\_4\_76

17. Vladzimirsky AV, Gusev AV, Sharova DE, et al. Health Information System Maturity Assessment Methodology. *Medical Doctor and IT*. 2022;(3):68–84. doi: 10.25881/18110193\_2022\_3\_68

18. Order of the Moscow City Health Department No. 160 dated 03.11.2022 «On Approval of the Procedure and Conditions for Conducting an Experiment on the Use of Innovative Technologies in Computer Vision for Analyzing Medical Images and Further Application in the Moscow City Health Care System in 2022». (In Russ).

## ОБ АВТОРАХ

\* **Зинченко Виктория Валерьевна;**

адрес: Российская Федерация, 127051, Москва, ул. Петровка, д. 24, стр. 1;

ORCID: 0000-0002-2307-725X;

eLibrary SPIN: 4188-0635;

e-mail: ZinchenkoVV1@zdrav.mos.ru

**Арзамасов Кирилл Михайлович,** канд. мед. наук;

ORCID: 0000-0001-7786-0349;

eLibrary SPIN: 3160-8062;

e-mail: ArzamasovKM@zdrav.mos.ru

**Кремнева Елена Игоревна,** канд. мед. наук;

ORCID: 0000-0001-9396-6063;

eLibrary SPIN: 8799-8092;

e-mail: KremnevaEI@zdrav.mos.ru

**Владзimirский Антон Вячеславович,** д-р мед. наук;

ORCID: 0000-0002-2990-7736;

eLibrary SPIN: 3602-7120;

e-mail: VladzimirskijAV@zdrav.mos.ru

**Васильев Юрий Александрович,** канд. мед. наук;

ORCID: 0000-0002-0208-5218;

eLibrary SPIN: 4458-5608;

e-mail: VasilevYA1@zdrav.mos.ru

## AUTHORS' INFO

\* **Viktoria V. Zinchenko;**

address: 24-1 Petrovka str., 1127051, Moscow, Russian Federation;

ORCID: 0000-0002-2307-725X;

eLibrary SPIN: 4188-0635;

e-mail: ZinchenkoVV1@zdrav.mos.ru

**Kirill M. Arzamasov,** MD, Cand. Sci. (Med.);

ORCID: 0000-0001-7786-0349;

eLibrary SPIN: 3160-8062;

e-mail: ArzamasovKM@zdrav.mos.ru

**Elena I. Kremneva,** MD, Cand. Sci. (Med.);

ORCID: 0000-0001-9396-6063;

eLibrary SPIN: 8799-8092;

e-mail: KremnevaEI@zdrav.mos.ru

**Anton V. Vladzimirskyy,** MD, Dr. Sci. (Med.);

ORCID: 0000-0002-2990-7736;

eLibrary SPIN: 3602-7120;

e-mail: VladzimirskijAV@zdrav.mos.ru

**Yuriy A. Vasilev,** MD, Cand. Sci. (Med.);

ORCID: 0000-0002-0208-5218;

eLibrary SPIN: 4458-5608;

e-mail: VasilevYA1@zdrav.mos.ru

\* Автор, ответственный за переписку / Corresponding author