

ТЫРОВ И.А.,

Департамент здравоохранения города Москвы, Москва, Россия,
e-mail: zdrav@mos.ru

ВАСИЛЬЕВ Ю.А.,

к.м.н., ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия,
e-mail: info@npcmr.ru

АРЗАМАСОВ К.М.,

к.м.н., ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия,
e-mail: ArzamasovKM@zdrav.mos.ru

ВЛАДИМИРСКИЙ А.В.,

д.м.н., профессор, ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия,
e-mail: info@npcmr.ru

ШУЛЬКИН И.М.,

ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия,
e-mail: i.shulkin@npcmr.ru

ОМЕЛЯНСКАЯ О.В.,

ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия,
e-mail: info@npcmr.ru

ЧЕТВЕРИКОВ С.Ф.,

к.т.н., ГБУЗ «НПКЦ ДиТ ДЗМ», Москва, Россия,
e-mail: ChetverikovSF@zdrav.mos.ru

ОЦЕНКА ЗРЕЛОСТИ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ЗДРАВООХРАНЕНИЯ: МЕТОДОЛОГИЯ И ЕЕ ПРИМЕНЕНИЕ НА МАТЕРИАЛАХ МОСКОВСКОГО ЭКСПЕРИМЕНТА ПО КОМПЬЮТЕРНОМУ ЗРЕНИЮ В ЛУЧЕВОЙ ДИАГНОСТИКЕ

DOI: 10.25881/18110193_2022_4_76

Аннотация.

Цель работы разработать и апробировать методологию оценки зрелости программного обеспечения на основе технологии искусственного интеллекта (ТИИ) для сферы здравоохранения.

Материалы и методы. Методология разработки матрицы зрелости программного обеспечения на основе ТИИ для сферы здравоохранения основана на литературных данных и на анализе собственного практического опыта, полученного в ходе «Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы» в 2021–2022 гг. Изучены результаты работы 35 отдельных программных продуктов на основе ТИИ, охватывающих основные направления лучевой диагностики.

Результаты. Разработана матрица зрелости, учитывающая показатели технической стабильности — удельный вес технологических дефектов, и диагностическую составляющую — площадь под характеристической кривой. Данная модель апробирована на 35 программных продуктах на основе ТИИ. Зрелости достигли 40% рассмотренных программных продуктов. Для 24 программных продуктов на основе ТИИ проведена оценка динамики развития: 15 из них (62%) находятся в зоне диагностической стагнации; 8 (33%) — в зоне высокого диагностического и технического потенциала, 1 (4%) — в зоне низкого диагностического и технического потенциала и 1 (4%) при развитии диагностического потенциала ухудшил техническую составляющую.

Заключение. По результатам оценки качества работы 35 программных продуктов на основе ТИИ разработана методология оценки зрелости ТИИ для здравоохранения, которая включает в себя матрицу зрелости и метод оценки клинико-технической трансформации зрелости, что позволяет проводить оценку программного продукта на основе ТИИ как дискретно (одномоментно), так и в динамике.

Ключевые слова: лучевая диагностика; искусственный интеллект; оценка зрелости технологии.

Для цитирования: Тыров И.А., Васильев Ю.А., Арзамасов К.М., Владимирский А.В., Шулькин И.М., Омелянская О.В., Четвериков С.Ф. 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: 76-92. doi: 10.25881/18110193_2022_4_76.

TYROV I.A.,

Moscow Healthcare Department, Moscow, Russia,
e-mail: zdrav@mos.ru

VASILYEV Y.A.,

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, Moscow, Russia, e-mail: info@npcmr.ru

ARZAMASOV K.M.,

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, Moscow, Russia, e-mail: k.arzamasov@npcmr.ru

VLADZIMIRSKY A.V.,

DSc, Professor, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, Moscow, Russia, e-mail: info@npcmr.ru

SHULKIN I.M.,

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, Moscow, Russia, e-mail: i.shulkin@npcmr.ru

OMELYANSKAYA O.V.,

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, Moscow, Russia, info@npcmr.ru

CHETVERIKOV S.F.,

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies
of the Moscow Health Care Department, Moscow, Russia, e-mail: ChetverikovSF@zdrav.mos.ru

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

DOI: 10.25881/18110193_2022_4_76

Abstract.

Aim: to develop and test a methodology for assessing the maturity of healthcare software based on artificial intelligence (AI). Materials and methods. The methodology for developing a maturity matrix for AI-based healthcare software is based on published data and on an analysis of our own practical experience obtained during the «Experiment on the use of innovative technologies in the field of computer vision for the analysis of medical images and further application in the Moscow healthcare system» in 2021–2022. We studied study results from 35 separate software products based on AI, covering key areas of radiology.

Results. We developed a maturity matrix that takes into account the indicators of technical stability — the proportion of technological defects, and the diagnostic component — the area under the characteristic curve. This model has been tested in 35 software products based on AI, with 40% of the products having achieved maturity. The dynamics of development was assessed for 24 software products based on AI: 15 of them (62%) were in the zone of diagnostic stagnation; 8 (33%) — in the zone of high diagnostic and technical potential, 1 (4%) — in the zone of low diagnostic and technical potential, and 1 (4%) worsened the technical component with the increase in diagnostic potential.

Conclusion. A methodology for assessing the maturity of AI for healthcare has been developed based on the performance and quality assessment of 35 software products. This methodology includes a maturity matrix and a method for assessing the clinical and technical transformation of maturity, which makes it possible to evaluate an AI-based software product both discretely (simultaneously) and in dynamics.

Keywords: radiology; artificial intelligence; technology maturity assessment.

For citation: Tyrov I.A., Vasilyev Y.A., Arzamasov K.M., Vladzimirsky A.V., Shulkin I.M., Omelyanskaya O.V., Chetverikov S.F. 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 information technology.* 2022; 4: 76-92. doi: 10.25881/18110193_2022_4_76.

ВВЕДЕНИЕ

Информатизация и автоматизация современного здравоохранения приводят к масштабным трансформациям производственных процессов и сложившихся практик. Применение разнообразных информационных и телекоммуникационных технологий носит сквозной характер, охватывая все уровни и этапы оказания медицинской помощи, все клинические дисциплины и направления. В связи с этим все более актуальной становится оценка результативности и эффективности цифровизации. Ведь каждое технологическое решение в медицине — это лишь инструмент, использование которого требует обоснованного целеполагания, определения возможностей и ограничений, установления измеримых результатов. В последние годы появилось понятие «цифровая зрелость», фигурирующее в научных и методических материалах, а в Российской Федерации и некоторых иных странах введенное в нормативно-правовые документы [1–4]. В частности, оценка эффективности деятельности высших должностных лиц и органов исполнительной власти субъектов Российской Федерации предполагает в том числе расчет показателя цифровой зрелости (соответствующая методика утверждена Постановлением Правительства РФ от 03.04.2021 №542). Именно для области медицины разработан индекс цифровой зрелости здравоохранения, включенный в качестве инструмента в нормативно-правовые акты [5]. Созданы инструменты для оценки цифровой зрелости систем здравоохранения (например, «Global Digital Health Index»), медицинских организаций («Healthcare Information and Management Systems Society (HIMSS)» и т.д.) и даже отдельных информационных систем («Health Information Systems Interoperability Maturity Toolkit» и т.д.) [6–8].

В практическом аспекте разнообразные индексы и методики применяются для бенчмаркинга и создания рейтингов медицинских организаций или систем здравоохранения административно-территориальных единиц, а также для поддержки принятия и мониторинга выполнения управленческих решений и в научных задачах. В России, как уже было сказано выше, индекс цифровой зрелости здравоохранения применяется для оценки качества управленческой деятельности и развития системы медицинской помощи [9–12].

Большинство инструментов и методик предполагают в той или иной мере обобщенную оценку. Вместе с тем есть дискретные решения, то есть предназначенные для цифровизации отдельных клинических дисциплин и направлений, например, в сфере лучевой диагностики («Digital Imaging Adoption Model (DIAM)», «Медвиз») [13; 14]. Также в последние годы наметилась тенденция оценки цифровой зрелости с позиций пациентоориентированности соответствующих технологических решений и создаваемой на их основе системе медицинских услуг [15; 16].

В последние годы происходит стремительное развитие искусственного интеллекта (ИИ) — относительно нового поколения технологических решений для автоматизации процессов и отдельных задач в разных отраслях. В сфере здравоохранения технологии искусственного интеллекта (ТИИ) проходят этап активного научного изучения. В арсенале современного врача появляется все больше медицинских изделий на их основе. Активно формируется рынок соответствующего программного обеспечения (ПО).

Темпы развития медицинских ТИИ требуют создания новых специальных подходов к оценке их применимости, качества и результативности [17]. Вместе с тем только некоторые существующие инструменты оценки цифровой зрелости включают аспекты автоматизации (исключительно как компонент, по принципу «есть/нет» применение ТИИ). Такая ситуация затрудняет принятие управленческих решений, делает их «непрозрачными» из-за отсутствия объективных критериев для сравнения медицинских изделий на основе ТИИ.

В связи с этим нами была определена **цель исследования** — разработать и апробировать методологию оценки зрелости ПО на основе ТИИ для сферы здравоохранения.

МАТЕРИАЛ И МЕТОДЫ

Дизайн: исследование со смешанными методами, состоящее из качественного и количественного этапов с равными статусами [18].

Качественный этап исследования включал разработку матрицы зрелости ПО на основе ТИИ для сферы здравоохранения с учетом литературных данных (теоретический компонент) и собственного практического опыта (эмпирический

компонент). Используются аналитические методы научного познания: анализ, индукция, синтез. Количественный этап исследования включал определение уровня зрелости ПО на основе ТИИ, принимающего участие в научном «Эксперименте по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы» (далее — Московском Эксперименте). Используются методы описательной статистики.

Данные для исследования получены в ходе Московского Эксперимента в период с 2021 по 2022 гг. [19]. Эксперимент проводится в соответствии с Постановлением Правительства Москвы [20]. В указанный период времени в Московском Эксперименте приняло участие 14 компаний-разработчиков, представивших для участия 35 отдельных программных продуктов (сервисов на основе ИИ, далее — ИИ-сервисов). По специально разработанной нами процедуре и методологии ИИ-сервисы проходили этапное тестирование, интегрировались с ЕРИС ЕМИАС (Единый радиологический информационный сервис Единая медицинская информационно-аналитическая система) и начинали работать с потоком результатов лучевых исследований. Выполнялся автоматизированный анализ результатов по направлениям: компьютерная томография и/или низкодозная компьютерная томография органов грудной клетки с целью диагностики злокачественных новообразований легких (условное обозначение: ИИ1(КТ)-ИИ4(КТ)), изменений в легких при COVID-19 (ИИ1-ИИ7), компрессионных переломов тел позвонков (ИИ_ost_1, ИИ_ost_2), ишемической болезни сердца (коронарный кальций, паракардиальный жир) (ИИ_calc, ИИ_agat), аневризмы грудного отдела аорты с определением диаметра грудной аорты (ИИ_aort_1, ИИ_aort_2) расширения легочного ствола с определением диаметра легочного ствола (ИИ_trun_1, ИИ_trun_2), свободной жидкости (выпота) в плевральных полостях (ИИ_effus_1, ИИ_effus_2); компьютерная томография головного мозга для диагностики ишемического инсульта и внутримозговых кровоизлияний (ИИ_brain); компьютерная томография органов брюшной полости с целью диагностики компрессионных переломов тел

позвонков (ИИ_ost_abd), аневризмы брюшного отдела аорты с определением диаметра брюшной аорты (ИИ_aort_abd); маммография с целью диагностики рака молочной железы (ИИ1(ММГ) и ИИ2(ММГ)); рентгенография органов грудной клетки с целью диагностики различных патологий (ИИ1(РГ) — ИИ6(РГ)); флюорография легких с целью диагностики различных патологий (ИИ1(ФЛГ) — ИИ3(ФЛГ)).

В ходе анализа каждый ИИ-сервис должен решить три функциональные задачи: 1) сортировка результатов исследований в рабочем списке; 2) маркировка патологических находок на изображениях; 3) подготовка шаблона описания. Требования к структуре и содержанию результатов анализа содержатся в специально разработанных документах — базовых функциональных и диагностических требованиях. Результаты работы ИИ-сервисов становятся доступными на автоматизированных рабочих местах врачей-рентгенологов и представляют собой серию изображений с маркировкой и/или температурной картой, локализирующей находку, и текстовую информацию в виде краткого руководства пользователя, заключения и детализации по находкам. Врач использует или не использует результаты работы ИИ-сервисов по своему усмотрению, в том числе исходя из клинической ситуации. На этапе эксплуатации ИИ-сервисы подвергаются технологическому и клиническому мониторингу с целью контроля качества. Соответствующие оригинальные методологии были также разработаны нами [19]; часть из них вошла в состав национальных стандартов РФ по применению систем ИИ в медицине [21].

На момент подготовки данной статьи ИИ-сервисами в рамках Московского Эксперимента проанализированы результаты более 6 млн. лучевых исследований. 22 программных продукта получили статус медицинского изделия. Актуальная информация представлена на официальном сайте Московского Эксперимента — www.mosmed.ai.

В ходе мониторинга работы ИИ-сервисов определяются следующие параметры: процент технологических дефектов (несоответствия требованиям [22] к результатам работы ИИ-сервиса, которые приводят к потере клинической и диагностической ценности применения ИИ для анализа медицинских изображений),

проспективный ROCAUC (диагностическая точность по результатам проспективного сравнения заключения ИИ-сервиса с текстами протоколов-заключений врачей-рентгенологов согласно базовым диагностическим требованиям) и экспертная клиническая оценка (ручной просмотр ограниченного числа исследований с целью оценки правильности с диагностической точки зрения текстового заключения от ИИ, а также корректность локализации обнаруженных патологических находок ИИ-сервисом). Эти объективные, измеримые параметры были использованы нами в качестве базовых для создания матрицы зрелости.

РЕЗУЛЬТАТЫ

Для создания инструмента оценки зрелости ПО на основе ТИИ для здравоохранения была построена «матрица зрелости». В основу матрицы положены два параметра, отражающие успешность эксплуатации ИИ-сервисов: качество (совокупность свойств, существенных для использования по назначению) и эффективность (степень соответствия результатов работы, характеризующая приспособленность к достижению цели).

Для качественной составляющей матрицы была использована четырехпольная таблица в координатных осях (Рис. 1), где:

- Ось OX — процент технологических дефектов;
- Ось OY — единица минус проспективный ROCAUC;

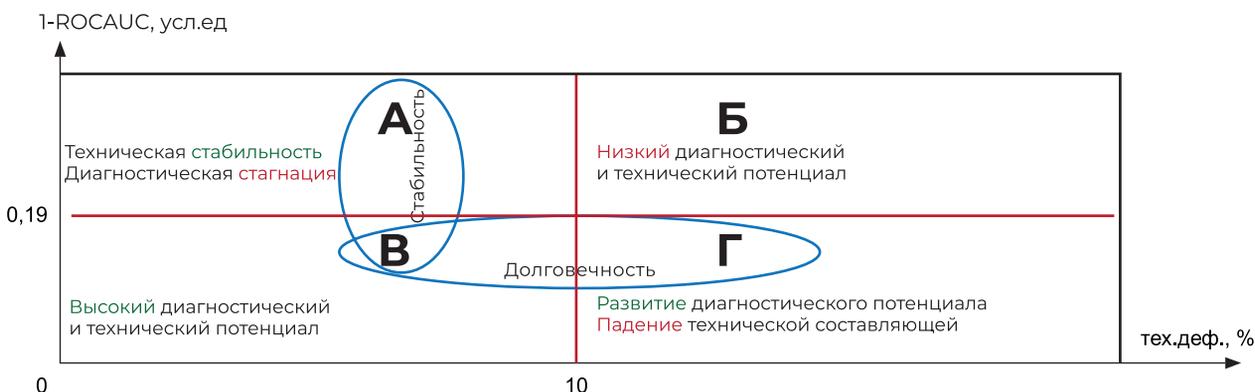


Рисунок 1 — Качественная составляющая матрицы. По горизонтальной оси — процент технологических дефектов; красная черта соответствует пороговому значению в 10%. По вертикальной оси — величина, обратная значению проспективного AUC; красная черта соответствует пороговому значению в 0,19. А, Б, В, Г — наименование зон развития ИИ-Сервиса, подробное пояснение в тексте.

- Граница «0,19» — горизонтальная линия с граничным значением для клинической значимости параметра «1 — ROCAUC» в соответствии с методическими рекомендациями [23], равное «1-0,81»;

- Граница «10» — вертикальная линия на уровне отметки в 10% технологических дефектов, в соответствии с Приказом [22].

Физический смысл матрицы зрелости ИИ-Сервиса заключается в выделении четырех категорий:

- зона «А», в которой ИИ-сервис достигает технической стабильности, но не развивает и не улучшает свою диагностическую составляющую;
- зона «Б», в которой ИИ-сервис не обладает на должном уровне техническим и диагностическим свойствами для осуществления качественной работы;
- зона «В», в которой ИИ-сервис обладает на должном уровне техническим и диагностическим свойствами для осуществления качественной работы;
- зона «Г», в которой ИИ-сервис развивает и улучшает свою диагностическую составляющую, но, как следствие, теряет свою техническую стабильность.

Качественная составляющая матрицы включает в себя два критерия:

- Стабильность — свойство ИИ-сервиса непрерывно сохранять свое качество при заданных воздействиях, характеризуется высоким техническим потенциалом (зоны А, В);

- Долговечность — свойство ИИ-сервиса сохранять свое качество при заданных воздействиях и условия восстановления свойств, характеризуется высоким диагностическим потенциалом (зоны В, Г).

Для эффективной составляющей матрицы была использована пузырьковая диаграмма (табл. 1), где:

- Диаметр пузырька — клиническая оценка работы ИИ-сервиса;
 - Орбита пузырька — разброс данных относительно среднего числа по клинической оценке.
- Эффективная составляющая матрицы включает в себя два критерия:
- Пригодность — соответствие ИИ-сервиса определенным клиничко-диагностическим требованиям (диаметр).

- Оптимальность — сбалансированная характеристика клиничко-диагностического параметра (орбита).

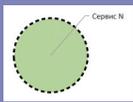
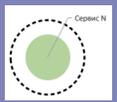
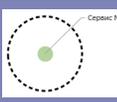
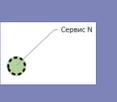
В зависимости от значений параметров возможны следующие варианты, которые представлены в таблице 1.

Матрица зрелости показывает клиническое качество результатов работы ИИ-сервисов в зависимости от клинической значимости и технологических дефектов (Рис. 2–5).

Матрица зрелости по сути является способом визуализации текущего положения ИИ-сервиса и может быть в перспективе использована для оценки развития (доработки) ИИ-сервиса.

Согласно предложенной матрице, наиболее зрелыми являются для компьютерной томографии COVID-19 ИИ-сервисы: ИИ 3, ИИ 2;

Таблица 1 — Варианты эффективной составляющей матрицы

Результат работы ИИ				
Варианты эффективной составляющей	Пригодный и оптимальный результат	Пригодный и неоптимальный результат	Непригодный и неоптимальный результат	Непригодный и оптимальный результат

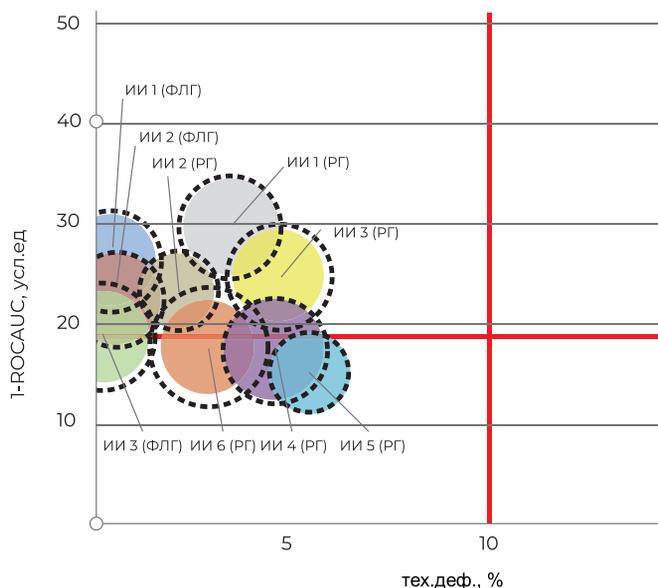


Рисунок 2 — Матрица по направлению рентгенография (РГ) органов грудной клетки, включая флюорографию (ФЛГ), ИИ (ФЛГ) — сервис на основе ТИИ по модальности флюорография, ИИ (РГ) — сервис на основе ТИИ по модальности рентгенография. По горизонтальной оси — процент технологических дефектов (тех.деф.,%); по вертикальной оси — величина «1-ROCAUC».

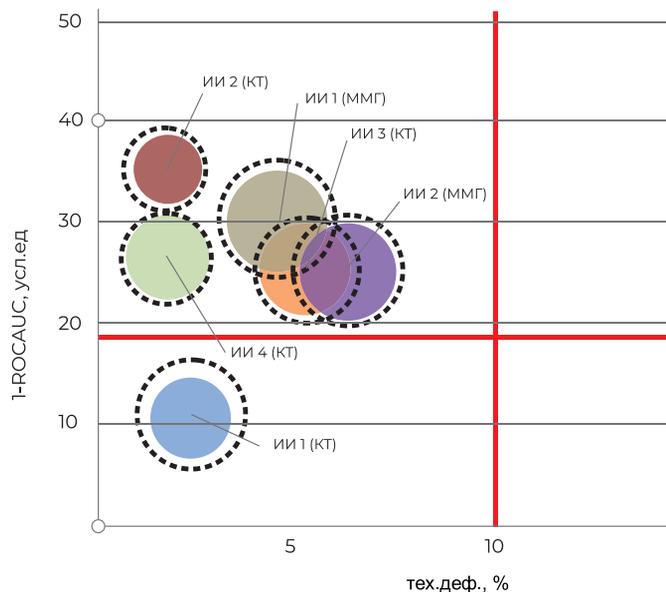


Рисунок 3. — Матрица по направлению компьютерная томография органов грудной клетки, целевая патология — рак лёгкого и маммография, целевая патология рак молочной железы. По горизонтальной оси — процент технологических дефектов (тех.деф.,%); по вертикальной оси — величина «1-ROCAUC».

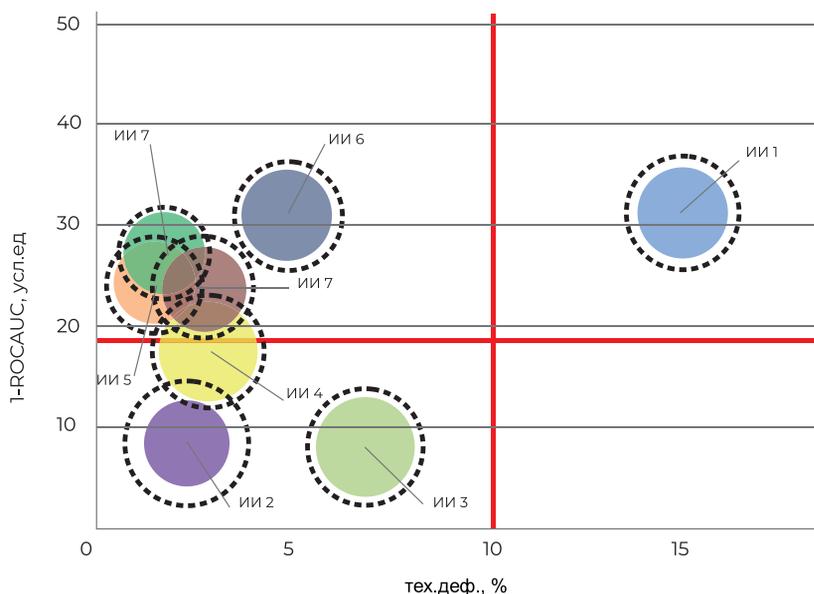


Рисунок 4 — Матрица по направлению компьютерная томография органов грудной клетки, целевая патология COVID-19. По горизонтальной оси — процент технологических дефектов (тех.деф.,%); по вертикальной оси — величина «1-ROCAUC».

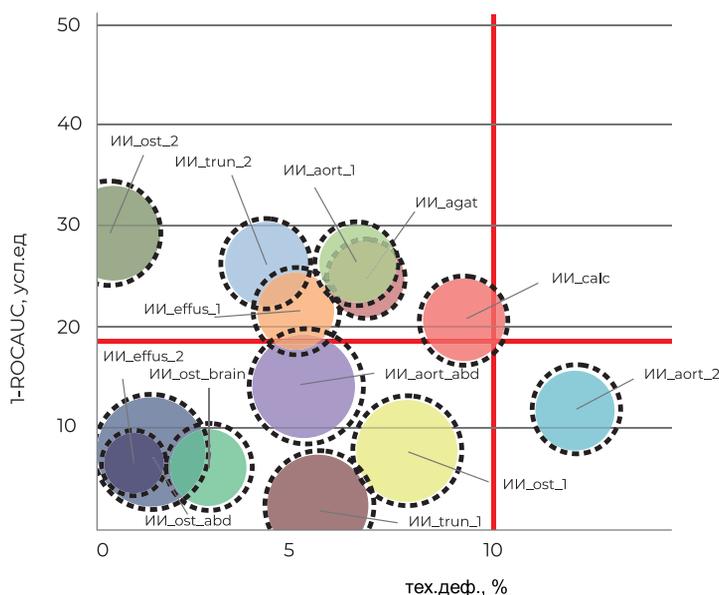


Рисунок 5 — Матрица по направлению компьютерная томография органов грудной клетки, брюшной полости и головного мозга. По горизонтальной оси — процент технологических дефектов (тех.деф.,%); по вертикальной оси — величина «I-ROCAUC».

для рентгенографии органов грудной клетки ИИ 6 (РГ), ИИ 5 (РГ), ИИ 4 (РГ) и ИИ 3 (ФЛГ); по другим направлениям компьютерной томографии ИИ_brain, ИИ_effus_2 и ИИ_aort_abd. Примечательно, что по модальности ММГ оба ИИ-сервиса показали удовлетворительные показатели технологической стабильности, но низкие метрики клинического качества работы.

Рентгенография органов грудной клетки: 3 ИИ-сервиса из 6 (50%) достигли своей зрелости и имеют показатели технологических дефектов ниже 10% и клиническую оценку более 0,81. 50% достигли технической стабильности и стремятся к своей клинической зрелости.

Флюорография легких: 1 ИИ-сервис из 3 (33%) достиг своей зрелости и имеет показатели технологических дефектов ниже 10% и клиническую оценку более 0,81. 66% достигли технической стабильности и стремятся к своей клинической зрелости.

Компьютерная томография органов грудной клетки с целью диагностики злокачественных новообразований легких: 1 ИИ-сервис из 4 (25%) достиг своей зрелости и имеет показатели технологических дефектов ниже 10% и клиническую

оценку более 0,81. 75% достигли технической стабильности и стремятся к своей клинической зрелости.

Компьютерная томография органов грудной клетки с целью диагностики COVID-19: 3 ИИ-сервиса из 8 (38%) достигли своей зрелости и имеют показатели технологических дефектов ниже 10% и клиническую оценку более 0,81. 50% достигли технической стабильности и стремятся к своей клинической зрелости. А также один ИИ-сервис успешно прошел входные тестирования, но при потоковой обработке исследований показал низкий уровень тех стабильности и клинической значимости (на момент написания статьи ИИ-сервис находился на доработке).

Маммография с целью диагностики рака молочной железы: 100% (2 из 2) достигли технической стабильности и стремятся к своей клинической зрелости.

Компьютерная томография органов грудной клетки/ брюшной полости: 6 ИИ-сервисов из 12 (50%) достигли своей зрелости и имеют показатели технологических дефектов ниже 10% и клиническую оценку более 0,81. 40% достигли технической стабильности и стремятся к своей

клинической зрелости. Также один ИИ-сервис при потоковой обработке исследований повысил клиническую значимость, но при осуществлении доработок потерял техническую стабильность.

Наряду с этим нельзя говорить об абсолютной зрелости ИИ-сервисов, т.к. полноценное, стабильное, точное и безошибочное его применение в рутинной практике врачей-рентгенологов требует стремления ИИ-сервиса в точку, соответствующую нулевому проценту технологических дефектов, а также к максимальной клинической оценке в 100%.

40% ИИ-сервисов достигли своей зрелости (14 из 35) по всем направлениям, что говорит о низком качестве предоставляемых решений. Для понимания ситуации и причин было

предложено использовать динамический показатель изменения параметров ИИ-сервисов.

В связи с тем, что ТИИ представляют собой достаточно динамично развивающиеся продукты, вызывает интерес оценка зрелости не только статичная, в данный момент времени, но и в динамике. Поэтому нами дополнительно разработана методика оценки клинико-технической трансформации зрелости ПО на основе ТИИ для здравоохранения.

Методика оценки клинико-технической трансформации заключается в отслеживании нахождения ИИ-сервиса в полях матрицы зрелости (см. выше) с учетом времени. На рисунках 6–9 приведены примеры развития ИИ-сервисов. Из 35 ранее проанализированных ИИ-сервисов было отобрано 24, которые имели минимум

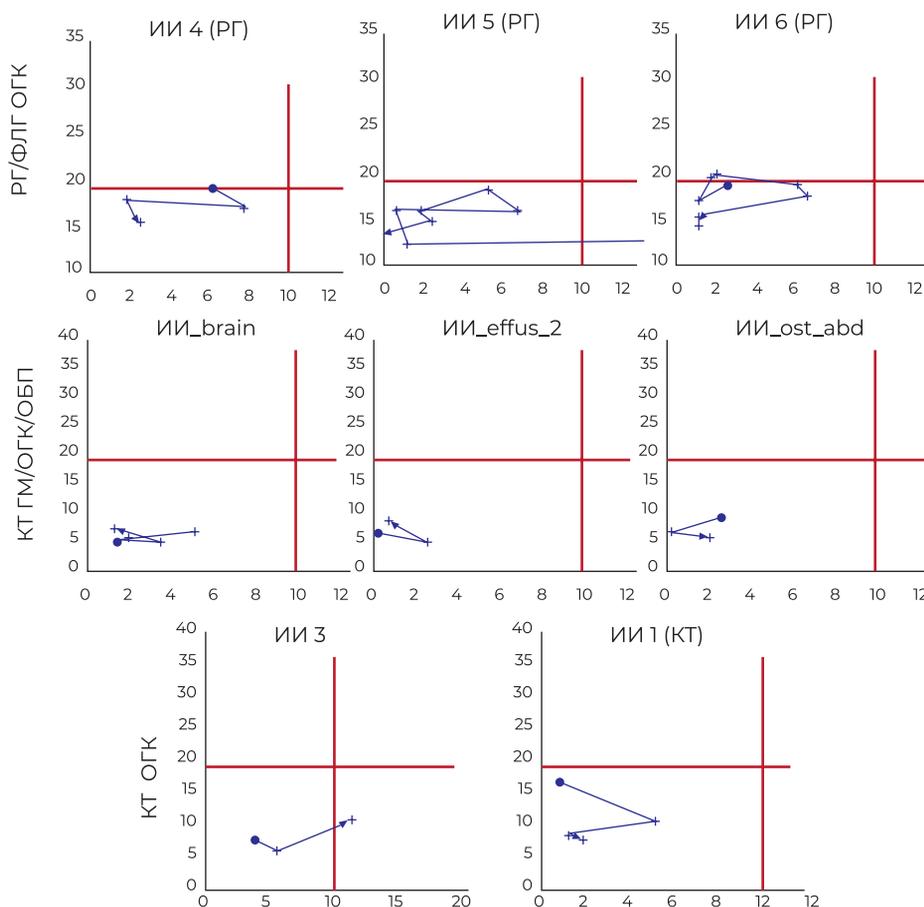


Рисунок 6 — Трансформация ИИ-сервисов, которые находятся в зоне высокого диагностического и технического потенциала.

3 точки для оценки динамики развития. Точкой на рисунке отмечено начало работы/аналитики ИИ-сервиса, стрелкой — завершение. Каждый крестик на прямой — это ежемесячная оценка ИИ-сервиса с учетом технической и клинической составляющей. Методика предоставления результата работы ИИ-сервиса в динамике

позволяет оценить тренд продукта ТИИ в целом, позволяет судить о направленности компаний-разработчиков на совершенствование ИИ-сервиса или наоборот. В качестве дополнительного примера на рисунке 9 приведен ИИ-сервис, для которого при развитии диагностического потенциала упала техническая составляющая.

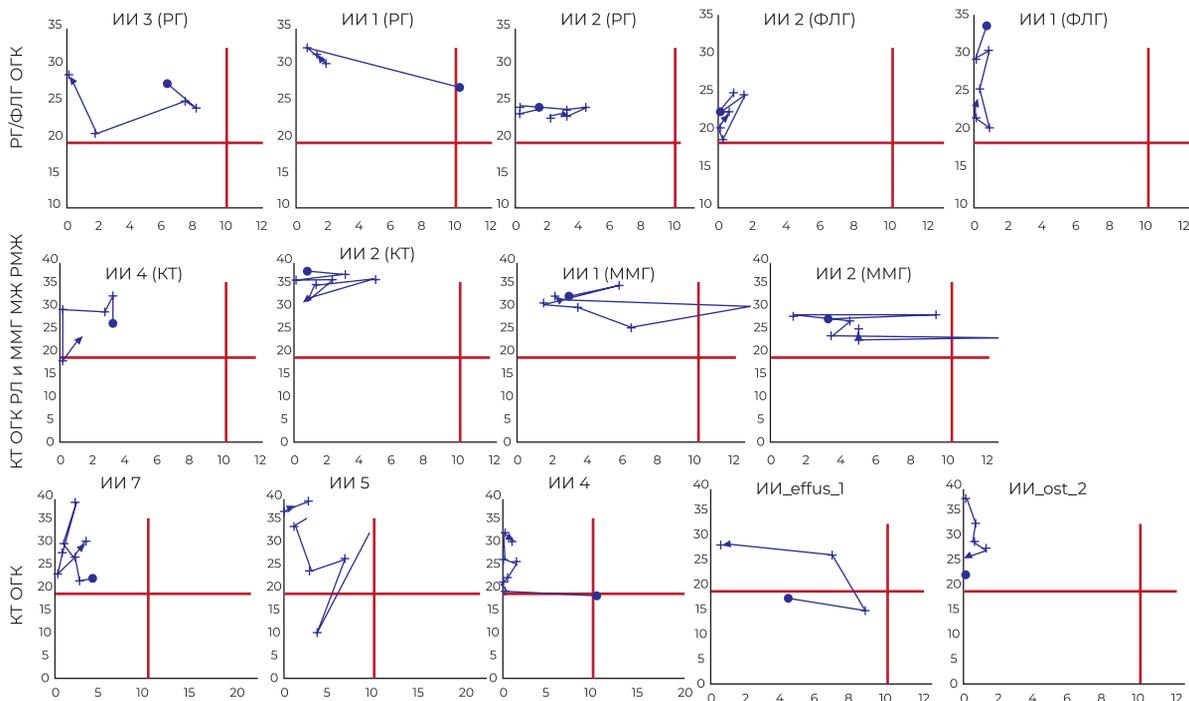


Рисунок 7 — Трансформация по направлению РГ/ФЛГ ОГК: находятся в зоне диагностической стагнации.

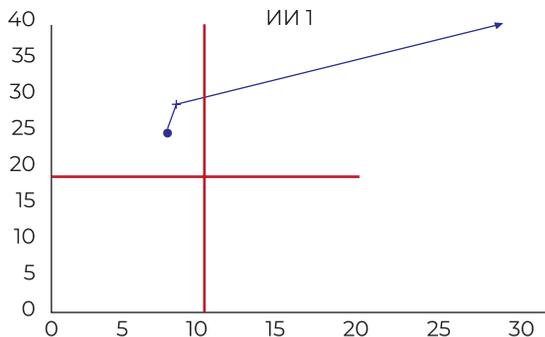


Рисунок 8 — Трансформация по направлению КТ ОГК COVID-19: находятся в зоне низкого диагностического и технического потенциала.

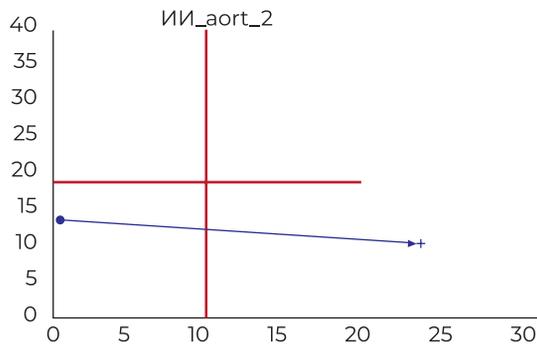


Рисунок 9 — Трансформация по другим направлениям КТ: при развитии диагностического потенциала упала техническая составляющая.

Оригинальная методика оценки клинико-технической трансформации применена для анализа зрелости ИИ-сервисов — участников Московского Эксперимента за 2022 год.

Согласно предложенной оценке трансформации присутствуют наиболее зрелые ИИ-сервисы в следующих направлениях: рентгенография патологии органов грудной клетки, компьютерная томография органов грудной клетки с целью диагностики COVID-19, компьютерная томография патологии органов брюшной полости, компьютерная томография патологии головного мозга.

Данная методика показывает, что для одного ИИ-сервиса (4%) при развитии диагностического потенциала упала техническая составляющая, один ИИ-сервис (4%) находятся в зоне низкого диагностического и технического потенциала; 15 ИИ-сервисов (62%) находятся в зоне диагностической стагнации; 8 ИИ-сервисов (33%) находятся в зоне высокого диагностического и технического потенциала.

Результаты работы ИИ-сервисов в ходе Московского Эксперимента указывают на возможность их применения в системе здравоохранения. По всем направлениям Эксперимента отмечается положительная динамика в части повышения технологической стабильности работы ИИ-сервисов. У отдельных направлений прослеживается положительная динамика клинической точности, но всё же требует дополнительной проработки ИИ-сервисов с целью повышения диагностического качества.

Положительная динамика в виде повышения диагностической точности в ходе проспективной работы ИИ-сервисов отмечается по всем существующим направлениям Эксперимента.

ОБСУЖДЕНИЕ

Разработка методик оценки зрелости ТИИ (для разных сфер и отраслей деятельности — ритейла, промышленности, транспорта) — это актуальная научно-практическая задача, что подтверждается количеством публикаций и материалами актуального систематического обзора [24]. Предложены различные подходы и инструменты оценки зрелости, впрочем, преимущественно сфокусированные на аспектах менеджмента, разработки, бизнес-стратегии, кадрового обеспечения, реже — качестве данных и самих алгоритмов [25–28].

На наш взгляд, имеет место выраженное смещение акцентов на оценку зрелости самой компании-разработчика, нежели на анализ качества и значимости продукта.

Как следует из данных систематического обзора только 46% исследований посвящены моделям оценки зрелости ИИ для конкретных отраслей [24]. На этом фоне ничтожно малое количество исследований посвящены проблематике зрелости ТИИ именно для сферы здравоохранения. Весьма общие положения изложены только в отношении области фармакологии (разработка новых лекарственных средств с применением ТИИ) [29; 30].

Предложенная нами матрица зрелости ИИ-сервисов представляет собой уникальное решение: за основу методологии положены не характеристики компании или процессы разработки, а качество и стабильность алгоритмов в условиях промышленной эксплуатации в рамках информационной системы в сфере здравоохранения. Для определения параметров качества и стабильности применяются специально разработанные методологии, учитывающие задачи и специфику автоматизации решения медицинских задач. Таким образом, нами предложен принципиально новый подход к оценке зрелости ТИИ в сфере здравоохранения.

В глобальной перспективе применимость и качество ТИИ в рентгенологии и радиологии изучены крайне поверхностно. Как следует из обзора клинических исследований существует значительное количество программных решений на основе ИИ для данной сферы медицины; все они имеют «широкий спектр зрелости и клинического использования». Однако существует большой пробел в изучении фактической эффективности инструментов ИИ в клинической практике [31]. По данным систематического обзора 535 статей установлено, что среднее количество пациентов, включаемых в исследования ТИИ в лучевой диагностике, составляет 460 [32]. На этом фоне Московский Эксперимент представляет собой крупнейшее в мире проспективное научное исследование применимости и результативности ТИИ в лучевой диагностике. Число участвующих ИИ-сервисов (65) и количество проанализированных исследований (7,8 млн.) значительно превосходят опубликованные результаты иных авторов [33–35].

Колоссальный объем первичных данных дает нам фундаментальную основу для теоретической и эмпирической разработки принципов, подходов, концепций, конкретных методологий. К таковой и относится матрица зрелости ИИ-сервисов.

В «медицинских» научных публикациях, как правило, исследуется точность, реже — результативность работы одного алгоритма. При этом стандартно проводится сравнение с аналогичными продуктами (чаще с литературными данными). В более «технических» статьях сопоставляется диагностическая точность нескольких вариантов алгоритмов и их комбинаций [36; 37]. Нами впервые проведен сопоставительный анализ качества и стабильности работы 35 независимых программных решений на основе ТИИ. Благодаря чему получены уникальные знания о применимости ПО на основе ИИ в лучевой диагностике.

Существует множество статей о результатах применения ИИ в лучевой диагностике, однако абсолютное большинство из них составляют ретроспективные когортные исследования (98%). Только в 14% исследований проводилась внешняя валидация. Из этого следует, что 86% публикаций о результативности ИИ в лучевой диагностике имеют высокий риск систематической ошибки. Необходимость внешней независимой валидации точности и качества ТИИ для здравоохранения утверждалась многим авторами. Цитируемый систематический обзор доказывает эту необходимость объективно: по итогам независимого тестирования метрики точности алгоритмов снижаются в среднем на 6% от заявленных разработчиками (диапазон снижения составляет 4–44%) [32].

В условиях Московского Эксперимента проводится внешняя независимая валидация ИИ-сервисов, результаты которой полностью подтверждают приведенные выше утверждения. Самооценка разработчиков часто носит завышенный, ярко выраженный рекламный характер, она явно требует той самой независимой внешней валидации [38–40].

Попытка же рейтингования или оценки зрелости ТИИ на основе коммерческих составляющих (уровня общественного резонанса, объема инвестиций и т.д.) [41] вообще недопустима для сферы здравоохранения, где на кону находятся безопасность и состояние здоровья человека.

Дополнительно подчеркнем, что дизайн Московского Эксперимента носит проспективный характер, который качественным образом отличает его от 98% иных исследований применимости ИИ в лучевой диагностике и позволяет получить уникальные объективные данные о точности ТИИ.

ВЫВОДЫ

На основе объективных результатов о работоспособности и качестве 35 ИИ-сервисов разработана методология оценки зрелости ТИИ для здравоохранения. Методология включает матрицу зрелости и метод оценки клинично-технической трансформации зрелости, что позволяет проводить оценку как дискретно (одномоментно), так и в динамике. Применение разработанной нами модели на практике демонстрирует достижение высокой степени зрелости для четверти из проанализированных решений на основе ТИИ (9 из 35), участвующих в Московском Эксперименте. Эти данные указывают на необходимость дальнейшего совершенствования ТИИ для лучевой диагностики.

ЛИТЕРАТУРА/REFERENCES

1. Пугачев П.С., Гусев А.В., Кобякова О.С., Кадыров Ф.Н., Гаврилов Д.В., Новицкий Р.Э., Владимирский А.В. Мировые тренды цифровой трансформации отрасли здравоохранения // Национальное здравоохранение. — 2021. — Т.2. — №2. — С.5-12. [Pugachev PS, Gusev AV, Kobjakova OS, Kadyrov FN, Gavrillov DV, Novickij Rje, Vladzimirskij AV. Global trends in the digital transformation of the healthcare industry. National Healthcare. 2021; 2(2): 5-12. (In Russ.)]
2. Лисицкий Н.Н., Антохин Ю.Н. Управление цифровым развитием организационной системы российского здравоохранения: национальная и глобальная повестка // Экономика. Право. Инновации. — 2021. — №4. — С.49-54. [Lisickij NN, Antohin JuN. Management of the digital development of the Russian healthcare organizational system: national and global agenda. Economy. Right. Innovation. 2021; 4: 49-54. (In Russ.)]
3. Гусев А.В., Владимирский А.В., Голубев Н.А., Зарубина Т.В. Информатизация здравоохранения Российской Федерации: история и результаты развития // Национальное здравоохранение. —

2021. — Т.2. — №3. — С.5-17.[Gusev AV, Vladzimirskij AV, Golubev NA, Zarubina TV. Informatization of healthcare in the Russian Federation: history and results of development. National Health. 2021; 2(3): 5-17. (In Russ.)]
4. Iyamu I, Xu АХТ, Gómez-Ramírez O, Ablona A, Chang HJ, Mckee G, Gilbert M. Defining Digital Public Health and the Role of Digitization, Digitalization, and Digital Transformation: Scoping Review. JMIR Public Health Surveill. 2021; 7(11): e30399. doi: 10.2196/30399.
 5. Орлов Г.М., Левин М.Б. Методологические подходы к разработке эталонных моделей государственных информационных систем в сфере здравоохранения субъектов Российской Федерации // Информационные ресурсы России. — 2021. — №2. — С.20-27. [Orlov GM, Levin MB. Methodological approaches to the development of reference models of state information systems in the field of health care of the constituent entities of the Russian Federation. Information resources of Russia. 2021; 2(180): 20-27. (In Russ.)]
 6. Cresswell K, Sheikh A, Krasuska M, Heeney C, Franklin BD, Lane W, Mozaffar H, Mason K, Eason S, Hinder S, Potts HWW, Williams R. Reconceptualising the digital maturity of health systems. Lancet Digit Health. 2019; 1(5): e200-e201. doi: 10.1016/S2589-7500(19)30083-4.
 7. Kouroubali A, Papastilianou A, Katehakis DG. Preliminary Assessment of the Interoperability Maturity of Healthcare Digital Services vs Public Services of Other Sectors. Stud Health Technol Inform. 2019; 264: 654-658. doi: 10.3233/SHT1190304.
 8. Liaw ST, Zhou R, Ansari S, Gao J. A digital health profile & maturity assessment toolkit: cocreation and testing in the Pacific Islands. J Am Med Inform Assoc. 2021; 28(3): 494-503. doi: 10.1093/jamia/ocaa255.
 9. Duncan R, Eden R, Woods L, Wong I, Sullivan C. Synthesizing Dimensions of Digital Maturity in Hospitals: Systematic Review. J Med Internet Res. 2022; 24(3): e32994. doi: 10.2196/32994.
 10. Woods L, Eden R, Pearce A, Wong YCI, Jayan L, Green D, McNeil K, Sullivan C. Evaluating Digital Health Capability at Scale Using the Digital Health Indicator. Appl Clin Inform. 2022; 13(5): 991-1001. doi: 10.1055/s-0042-1757554.
 11. Орлов Г.М. Метод измерения цифровой зрелости региональной системы записи к врачу на основе эталонной сервисной модели. International Journal of Open Information Technologies. — 2020. — Т.8. — №11. — С. 110-121. [Orlov GM. A method for measuring the digital maturity of a regional appointment system based on a reference service model. International Journal of Open Information Technologies. 2020; 8(11): 110-121. (In Russ.)]
 12. Шулькин И.М., Владимирский А.В. Управление на основе данных в лучевой диагностике: оценка результативности модели Единого радиологического информационного сервиса // Менеджер здравоохранения. — 2022. — №7. — С.68-80. [Shul'kin IM, Vladzimirskij AV. Data-driven management in radiology: assessment of the effectiveness of the Unified Radiological Information Service model. Health manager. 2022; 7: 68-80. (In Russ.)]
 13. Studzinski J. Evaluating the maturity of IT-supported clinical imaging and diagnosis using the Digital Imaging Adoption Model : Are your clinical imaging processes ready for the digital era? Radiologe. 2017; 57(6): 466-469. doi: 10.1007/s00117-017-0253-8.
 14. Морозов С.П., Владимирский А.В., Сафронов Д.С. Бенчмаркинг для оценки качества цифровизации отделений лучевой диагностики: разработка методологии // Врач и информационные технологии. — 2019. — №1. — С.40-45. [Morozov SP, Vladzimirskij AV, Safronov DS. Benchmarking for assessing the quality of digitization of radiology departments: methodology development Physician and information technology. 2019; 1: 40-45. (In Russ.)]
 15. Flott K, Callahan R, Darzi A, Mayer E. A Patient-Centered Framework for Evaluating Digital Maturity of Health Services: A Systematic Review. J Med Internet Res. 2016; 18(4): e75. doi: 10.2196/jmir.5047.
 16. Khanbhai M, Flott K, Darzi A, Mayer E. Evaluating Digital Maturity and Patient Acceptability of Real-Time Patient Experience Feedback Systems: Systematic Review. J Med Internet Res. 2019; 21(1): e9076. doi: 10.2196/jmir.9076.
 17. Гусев А.В., Астапенко Е.М., Иванов И.В., Зарубина Т.В., Кобринский Б.А. Принципы формирования доверия к системам искусственного интеллекта для сферы здравоохранения // Вестник Росздравнадзора. — 2022. — №2. — С.25-33.[Gusev AV, Astapenko EM, Ivanov IV, Zarubina TV,

- Kobrinskij BA. Principles of building trust in artificial intelligence systems for healthcare. *Bulletin of Roszdravnadzor*. 2022; 2: 25-33. (In Russ.)]
18. Schoonenboom J, Johnson RB. How to Construct a Mixed Methods Research Design. *Kolner Z Soz Sozpsychol*. 2017; 69(2): 107-131. doi:10.1007/s11577-017-0454-1.
 19. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента: монография / Под ред. Ю.А. Васильева, А.В. Владимировского. — М. Ридеро, 2022. — 388 с. [Computer vision in radiology: the first stage of the Moscow experiment: monograph. Y.A. Vasilyev, A.V. Vladimirovsky, editors. 2022. 388p (In Russ.)]
 20. Постановление Правительства Москвы от 21.11.2019 №1543-ПП (в ред. от 22.02.2022 №225-ПП). [Decree of the Government of Moscow 21.11.2019 г. № 1543-ПП (Ed. 22.02.2022 №225-ПП) (In Russ.)]
 21. Морозов С.П., Владимировский А.В., Шарова Д.Е., Ахмад Е.С., Зинченко В.В. Первые национальные стандарты Российской Федерации на системы искусственного интеллекта в медицине // Менеджмент качества в медицине. — 2022. — №1. — С.58-62. [Morozov SP, Vladymyrskyy AV, Sharova DE, Akhmad ES, Zinchenko VV. The first russian federation national standards for artificial intelligence systems in medicine. *Quality management in medicine*. 2022; 1: 58-62(In Russ.)]
 22. Приказ Департамента здравоохранения города Москвы от 24.02.2022 №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». [Order of the Moscow Department of Health of 24.02.2022 №160 «On approval of the Procedure and conditions for conducting an experiment on the use of innovative technologies in the field of computer vision for the analysis of medical images and further application in the healthcare system of the city of Moscow». (In Russ.)]
 23. Морозов С.П., Владимировский А.В., Кляшторный В.Г. и др. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика). — Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, 2019. [Morozov SP, Vladimirovskij AV, Kljashtornyj VG, et all. Clinical trials of software based on intelligent technologies (radiation diagnostics). Moscow: Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow City Health Department, 2019. (In Russ.)]
 24. Sadiq RB, Safie N, Abd Rahman AH, Goudarzi S. Artificial intelligence maturity model: a systematic literature review. *PeerJ Comput Sci*. 2021; 7: e661. doi: 10.7717/peerj-cs.661.
 25. Coates DL, Martin A. An instrument to evaluate the maturity of bias governance capability in artificial intelligence projects. *IBM Journal of Research and Development*. 2019; 4(5): 1-7; 15. doi: 10.1147/JRD.2019.2915062.
 26. Kreutzer RT, Sirrenberg M. AI challenge — how artificial intelligence can be anchored in a company. *Understanding Artificial Intelligence 2020*; Cham: Springer; 2020: 235-273.
 27. Lichtenthaler U. Five maturity levels of managing AI: from isolated ignorance to integrated intelligence. *Journal of Innovation Management*. 2020; 8(1): 39-50. doi: 10.24840/2183-0606_008.001_0005.
 28. Ellefsen APT, Oleśków-Szłapka J, Pawłowski G, Tobała A. Striving for excellence in AI implementation: AI maturity model framework and preliminary research results. *LogForum*. 2019; 15: 363-376.
 29. Tu H, Lin Z, Lee K. Automation With Intelligence in Drug Research. *Clin Ther*. 2019; 41(11): 2436-2444. doi: 10.1016/j.clinthera.2019.09.002.
 30. Lamberti MJ, Wilkinson M, Donzanti BA, Wohlhieter GE, Parikh S, Wilkins RG, Getz K. A Study on the Application and Use of Artificial Intelligence to Support Drug Development. *Clin Ther*. 2019; 41(8): 1414-1426. doi: 10.1016/j.clinthera.2019.05.018.
 31. Tariq A, Purkayastha S, Padmanaban GP, Krupinski E, Trivedi H, Banerjee I, Gichoya JW. Current Clinical Applications of Artificial Intelligence in Radiology and Their Best Supporting Evidence. *J Am Coll Radiol*. 2020; 17(11): 1371-1381. doi: 10.1016/j.jacr.2020.08.018.
 32. Kelly BS, Judge C, Bollard SM, Clifford SM, Healy GM, Aziz A, Mathur P, Islam S, Yeom KW, Lawlor A, Killeen RP. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). *Eur Radiol*. 2022 Apr 14. doi: 10.1007/s00330-022-08784-6.

33. Jin C, Chen W, Cao Y, Xu Z, Tan Z, Zhang X, Deng L, Zheng C, Zhou J, Shi H, Feng J. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat Commun.* 2020; 11(1): 5088. doi: 10.1038/s41467-020-18685-1.
34. Namiri NK, Lee J, Astuto B, Liu F, Shah R, Majumdar S, Pedoia V. Deep learning for large scale MRI-based morphological phenotyping of osteoarthritis. *Sci Rep.* 2021; 11(1): 10915. doi: 10.1038/s41598-021-90292-6.
35. Stemmer A, Shadmi R, Bregman-Amitai O, Chetrit D, Blagev D, Orlovsky M, Deutsch L, Elnekave E. Using machine learning algorithms to review computed tomography scans and assess risk for cardiovascular disease: Retrospective analysis from the National Lung Screening Trial (NLST). *PLoS One.* 2020; 15(8): e0236021. doi: 10.1371/journal.pone.0236021.
36. Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, Stewart M, Collins GS, Furniss D. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology.* 2022 Jul; 304(1): 50-62. doi: 10.1148/radiol.211785.
37. Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg.* 2020; 12(2): 156-164. doi: 10.1136/neurintsurg-2019-015135.
38. Гаврилов П.В., Гаврилова О.П., Смольникова У.А. Выявление периферических образований в легких с использованием программы автоматизированного анализа флюорографических изображений // *Лучевая диагностика и терапия.* — 2020. — №51. — С.77. [Gavrilov PV, Gavrilova OP, Smol'nikova UA. Identification of peripheral formations in the lungs using a program for automated analysis of fluorographic images. *Radiation diagnostics and therapy.* 2020; S1: 77. (In Russ.)]
39. Дрокин И.С., Еричева Е.В., Бухвалов О.Л., Пилюс П.С., Малыгина Т.С., Синицын В.Е. Опыт разработки и внедрения системы поиска онкологических образований с помощью искусственного интеллекта на примере рентгеновской компьютерной томографии легких // *Врач и информационные технологии.* — 2019. — №3. — С.48-57. [Drokin IS, Ericheva EV, Buhvalov OL, Piljus PS, Malygina TS, Sinicyn VE. Experience in the development and implementation of a system for searching for oncological formations using artificial intelligence on the example of X-ray computed tomography of the lungs. *Physician and information technology.* 2019; 3: 48-57. (In Russ.)]
40. Павлович П.И., Броннов О.Ю., Капнинский А.А., Абович Ю.А., Рычагова Н.И. Сравнительное исследование результатов анализа данных цифровой маммографии системы на основе искусственного интеллекта «Цельс» и врачей-рентгенологов // *Digital Diagnostics.* — 2021. — Т.2. — №S2. — С.22-23. [Pavlovich PI, Bronov OJu, Kapninskij AA, Abovich JuA, Rychagova NI. Comparative study of the results of digital mammography data analysis of the system based on artificial intelligence «Celsus» and radiologists. *Digital Diagnostics.* 2021; 2(S2): 22-23. (In Russ.)]
41. Комарь П.А., Дмитриев В.С., Ледеяева А.М., Шадеркин И.А., Зеленский М.М. Рейтинг стартапов искусственного интеллекта: перспективы для здравоохранения России // *Российский журнал телемедицины и электронного здравоохранения.* — 2021. — Т.7. — №3. — С.32-41. [Komar' PA, Dmitriev VS, Ledjaeva AM, Shaderkin IA, Zelenskij MM. Rating of artificial intelligence startups: prospects for healthcare in Russia. *Russian journal of telemedicine and e-health.* 2021; 7(3): 32-41. (In Russ.)]