CENTER FOR DIAGNOSTICS AND TELEMEDICINE

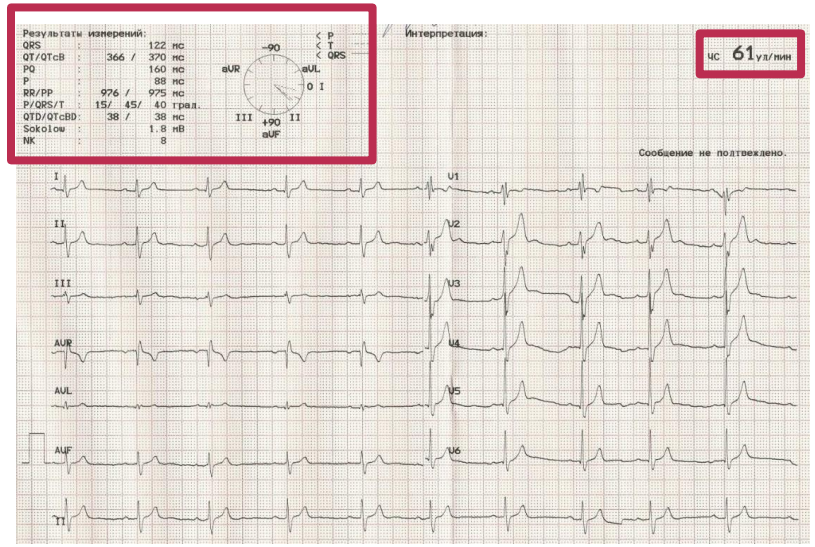
MOSCOW HEALTHCARE

AI Technologies in Healthcare: Infrastructure, Data, Reliability and Quality Assessment, Implementation (Experience and Evolution of the Moscow Experiment)

Anton Vladzymyrskyy
Ph.D. in Medicine, D.Sc., Deputy Director for R&D

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department. Moscow, 2025

Automation (or AI?) in medicine

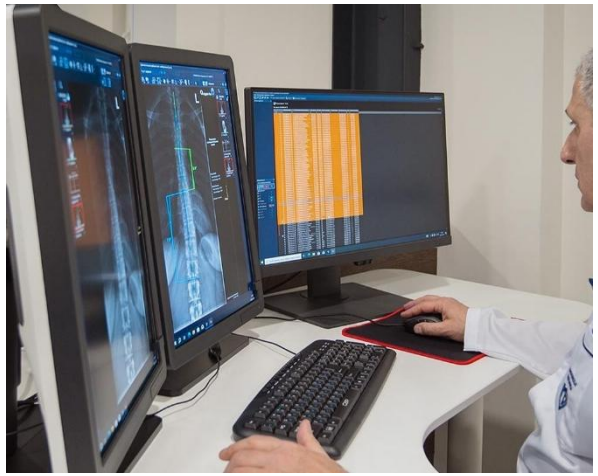


Electrocardiography

1960-1970.

Lab tests

>2000



Medical Imaging and EHR

Currently

Radiology: Current Challenges

The increasing demand for diagnostic imaging necessitates greater productivity and quality

Simply hiring more staff is costly, unsustainable for other specialties, and fails to address diagnostic accuracy

- Need for subspecialization – General radiologists lack the expertise to ensure consistent diagnostic quality;
- High workloads limit opportunities for continuous professional development;
- Workload distribution is uneven across health facilities, varying by patient volume and staff qualifications;
- medical scanners downtime due to radiologist shortages

Current resource allocation fails to match demand growth

Inefficient equipment use and incomplete diagnostic service offerings.

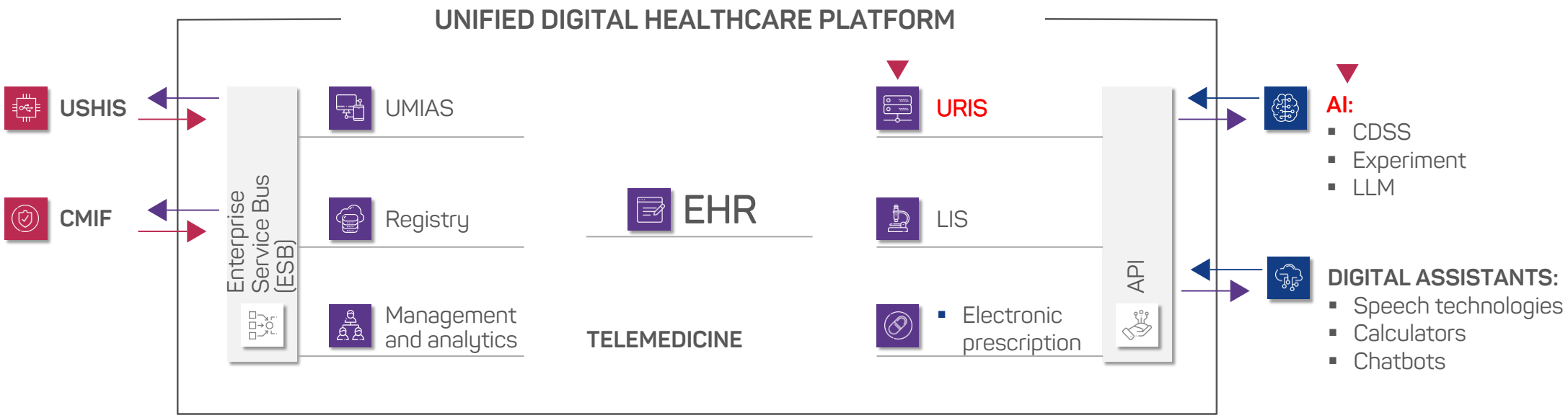
Report turnaround times often exceed the mandated 24-hour limit.

Patient access to diagnostic services remains **limited**

TELEMEDICINE+AI

Unified Digital Platform for Moscow Healthcare

The Unified Digital Platform is being implemented by the Department of Information Technology as part of the modernization of the Moscow social development complex



DOCTOR

- All medical records in the electronic format
- CDSS
- Doctor-to-Doctor Telemedicine
- Digital assistants

PATIENT

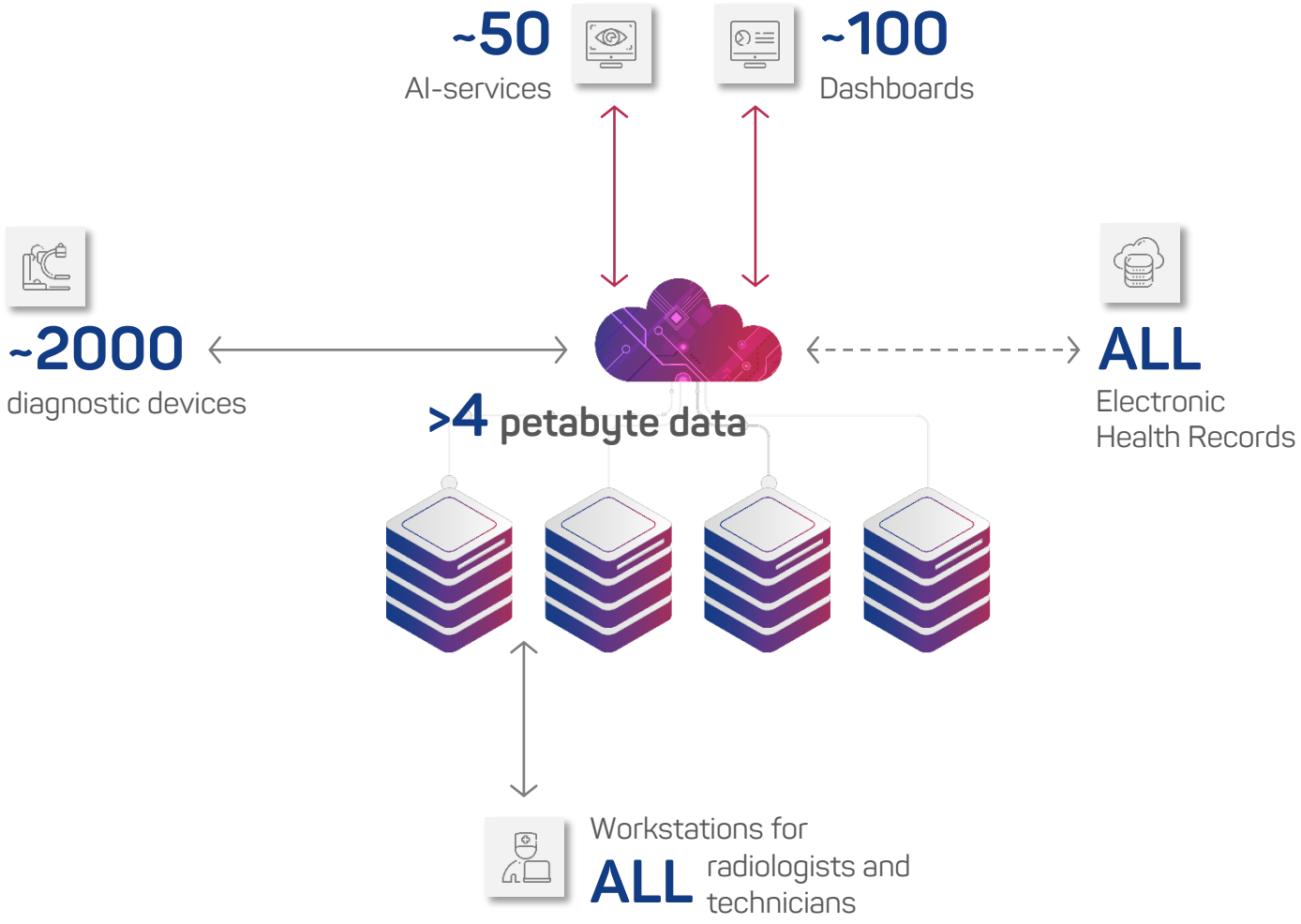
- All medical records in the electronic format on the mos.ru portal
- Mobile App "UMIAS.Info"
- Unified Digital Archive of Studies
- Electronic prescriptions
- Doctor-to-Patient Telemedicine

MEDICAL FACILITY

- Secure digital circuit for all clinics and hospitals
- Unified documentation archive
- Convenient monitoring tool
- Objective workload indicators

Unified Radiological Information Service (URIS)

telemedai.ru



Angiography	292 615
CT	9 789 996
Mammography	4 543 348
MRI	2 310 330
PET-CT	272 052
X-ray	32 623 884
Nuclear	140 838
Total	**49 973 063**

Independent opinion of KLAS Research:

- Globally, up to **300,000** radiology results annually are stored in **29.0%** of archives, and up to **100,000** – in **18.0%** of archives
- URIS is one of the **largest centralized repositories** in the global perspective
- Since 2020, URIS has been among the top **1.0%** of archives that accumulate **3 million** or more results of radiological studies annually

Artificial Intelligence in Medicine: Key Challenges

- **Clear task formulation**
- **Data set quality**
- **Independent external validation**
- **Automation** of real-world manufacturing process
- **Seamless Integration**
- **Validation** through prospective multicenter clinical trials
- **Explainability and reproducibility of results**
- **Lack of evaluation methodologies**

Introduction of Artificial Intelligence in Radiology in Moscow

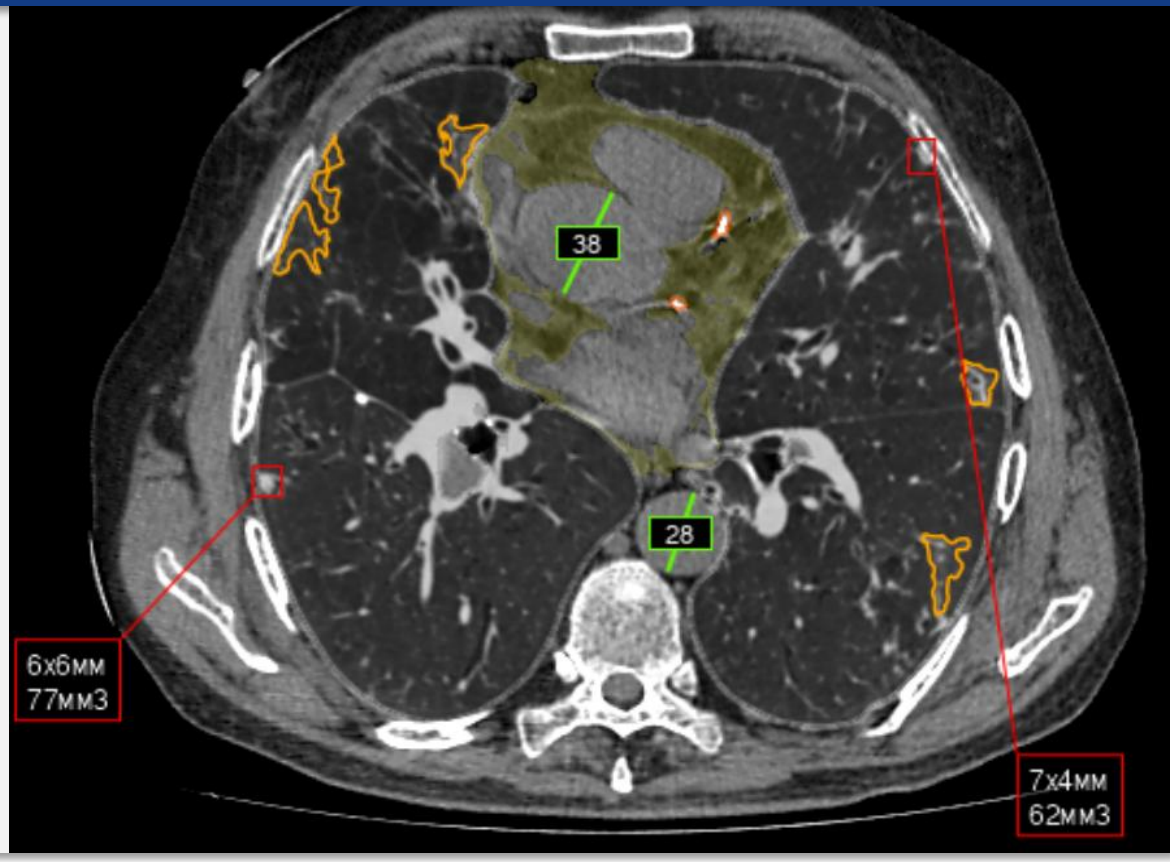
EXPERIMENT OF THE USE OF INNOVATIVE TECHNOLOGIES IN COMPUTER VISION for the analysis of medical images and further application in the Moscow healthcare system

OBJECTIVES:

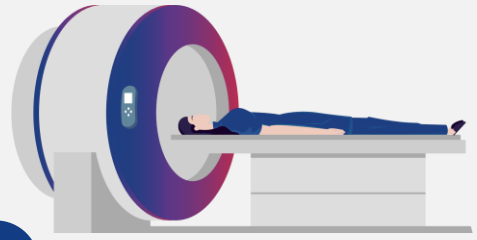
- a prospective multicenter study on the safety, applicability, and quality of artificial intelligence in radiology;
- establishing a market for artificial intelligence solutions in healthcare;
- digital transformation of healthcare.

PARTICIPANTS:

- Moscow Healthcare Department;
- Moscow Department of Information Technologies;
- Moscow medical organization;
- Center for Diagnostic and Telemedicine;
- AI software development companies.

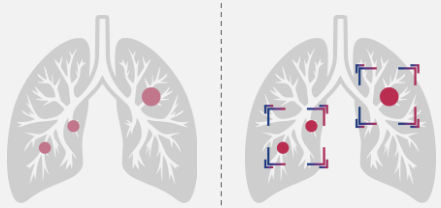


AI operation in Moscow radiology



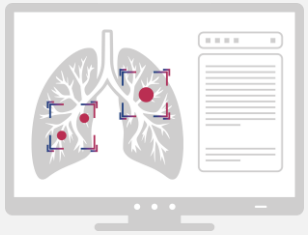
1 A patient is undergoing CT scan, MRI, or X-ray

2 All images are sent to the archive of the Moscow Health Data Processing Center

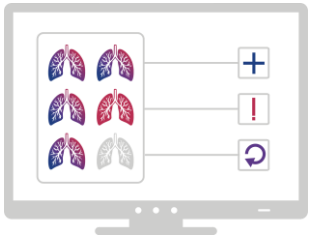


3 Original image

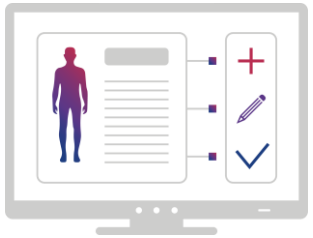
Image processed by AI



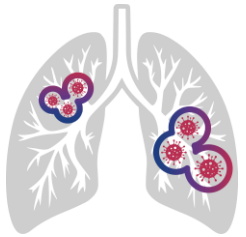
4 AI gives a hint to the radiologist



1 Automatic technical audit of the study



2 AI-assisted triage of a reporting worklist

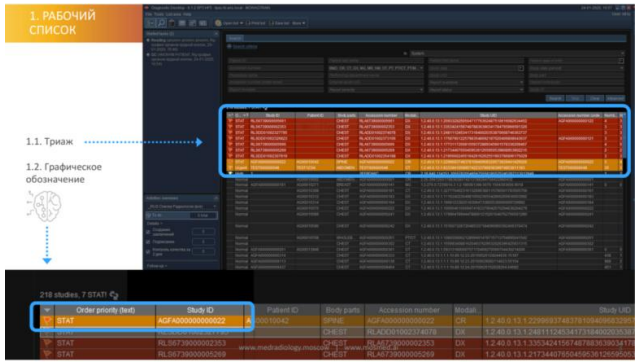


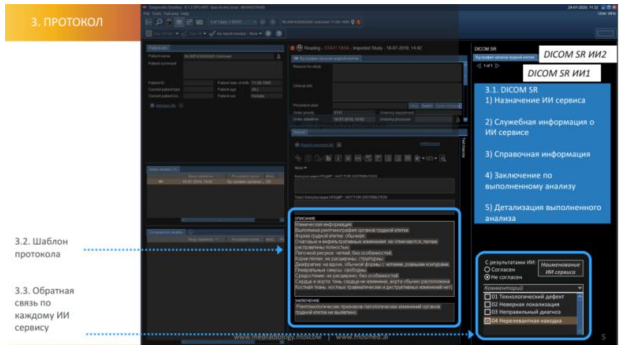
3 Graphical labeling of findings and measurements

4 AI report template

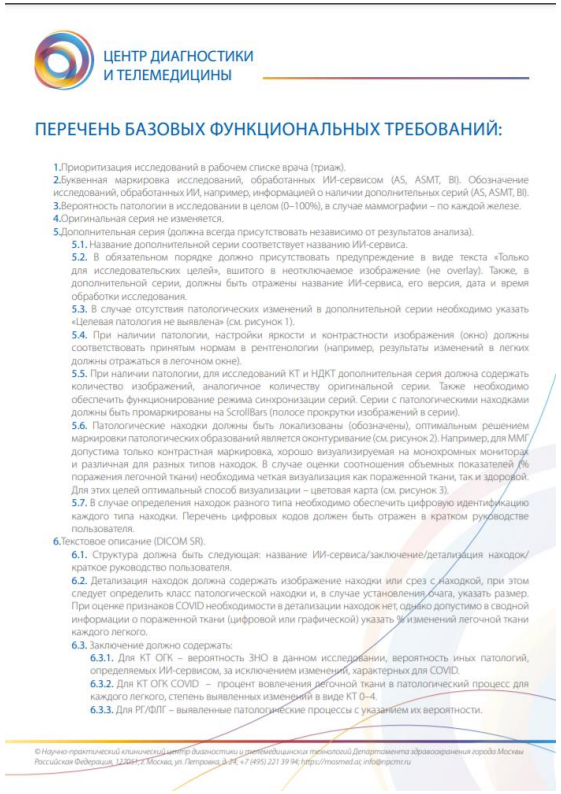
Standardization of technical requirements

February 2020





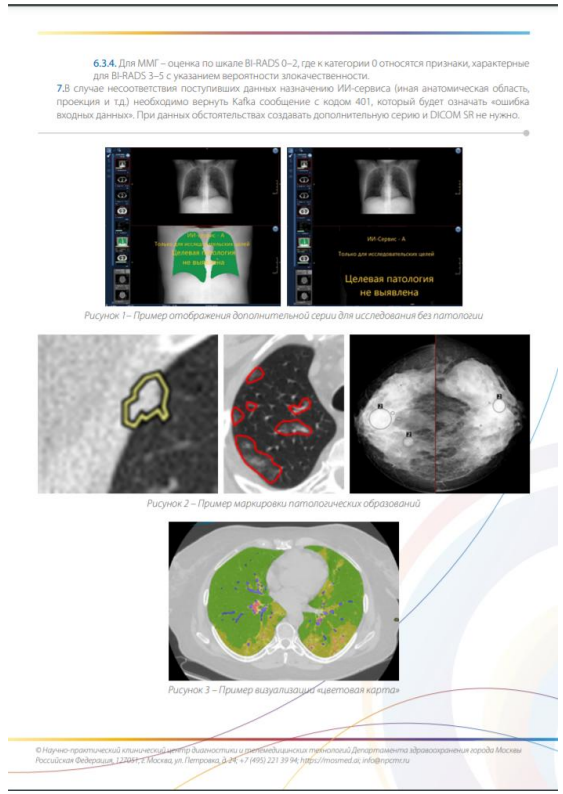
October 2024



ЦЕНТР ДИАГНОСТИКИ И ТЕЛЕМЕДИЦИНЫ

ПЕРЕЧЕНЬ БАЗОВЫХ ФУНКЦИОНАЛЬНЫХ ТРЕБОВАНИЙ:

1. Приоритизация исследований в рабочем списке врача (триаж).
2. Буквенная маркировка исследований, обработанных ИИ-сервисом (AS, ASMT, BI). Обозначение исследований, обработанных ИИ, например, информацией о наличии дополнительных серий (AS, ASMT, BI).
3. Вероятность патологии в исследовании в целом (0–100%), в случае маммографии – по каждой железе.
4. Оригинальная серия не изменяется.
5. Дополнительная серия (должна всегда присутствовать независимо от результатов анализа).
   - 5.1. Название дополнительной серии соответствует названию ИИ-сервиса.
   - 5.2. В обязательном порядке должно присутствовать предупреждение в виде текста «Только для исследовательских целей», вшитого в неотключаемое изображение (не overlay). Также, в дополнительной серии, должны быть отражены название ИИ-сервиса, его версия, дата и время обработки исследования.
   - 5.3. В случае отсутствия патологических изменений в дополнительной серии необходимо указать «Целевая патология не выявлена» (см. рисунок 1).
   - 5.4. При наличии патологии, настройки яркости и контрастности изображения (окно) должны соответствовать принятым нормам в рентгенологии (например, результаты изменений в легких должны отражаться в легочном окне).
   - 5.5. При наличии патологии, для исследований КТ и НДКТ дополнительная серия должна содержать количество изображений, аналогичное количеству оригинальной серии. Также необходимо обеспечить функционирование режима синхронизации серий. Серии с патологическими находками должны быть промаркированы на ScrollBars (полосе прокрутки изображений в серии).
   - 5.6. Патологические находки должны быть локализованы (обозначены), оптимальным решением маркировки патологических образований является оконтуривание (см. рисунок 2). Например, для ММГ допустима только контрастная маркировка, хорошо визуализируемая на монохромных мониторах и различная для разных типов находок. В случае оценки соотношения объемных показателей (% поражения легочной ткани) необходима четкая визуализация как пораженной ткани, так и здоровой. Для этих целей оптимальный способ визуализации – цветовая карта (см. рисунок 3).
   - 5.7. В случае определения находок разного типа необходимо обеспечить цифровую идентификацию каждого типа находки. Перечень цифровых кодов должен быть отражен в кратком руководстве пользователя.
6. Текстовое описание (DICOM SR).
   - 6.1. Структура должна быть следующая: название ИИ-сервиса/заключение/детализация находок/краткое руководство пользователя.
   - 6.2. Детализация находок должна содержать изображение находки или срез с находкой, при этом следует определить класс патологической находки и, в случае установления очага, указать размер. При оценке признаков COVID необходимости в детализации находок нет, однако допустимо в сводной информации о пораженной ткани (цифровой или графической) указать % изменений легочной ткани каждого легкого.
   - 6.3. Заключение должно содержать:
     - 6.3.1. Для КТ ОГК – вероятность ЗНО в данном исследовании, вероятность иных патологий, определяемых ИИ-сервисом, за исключением изменений, характерных для COVID.
     - 6.3.2. Для КТ ОГК COVID – процент вовлечения легочной ткани в патологический процесс для каждого легкого, степень выявленных изменений в виде КТ 0–4.
     - 6.3.3. Для РГ/ФЛГ – выявленные патологические процессы с указанием их вероятности.



6.3.4. Для ММГ – оценка по шкале BI-RADS 0–2, где к категории 0 относятся признаки, характерные для BI-RADS 3–5 с указанием вероятности злокачественности.
7. В случае несоответствия поступивших данных назначению ИИ-сервиса (иная анатомическая область, проекция и т.д.) необходимо вернуть Kafka сообщение с кодом 401, который будет означать «ошибка входных данных». При данных обстоятельствах создавать дополнительную серию и DICOM SR не нужно.

Рисунок 1 – Пример отображения дополнительной серии для исследования без патологии

Рисунок 2 – Пример маркировки патологических образований

Рисунок 3 – Пример визуализации «цветовая карта»

The basic functional requirements were modified based on the feedback from radiologists **to reduce the workload of radiologists and to standardize it.**

Standardization of medical requirements

Basic diagnostic requirements: clinical requirements for the AI service results when analyzing a specific type of study (1)
Answer the question: what signs of pathology an AI service should be able to identify, how to name them, and what classifications to use (2)

Forming:

- Developed by experts in the specialty
- Based on: clinical guidelines, approved classifications, and publications (3)
- Approved by the Scientific Committee on Problems

1 →
2 →
3 →

Базовые диагностические требования к результатам работы ИИ-сервисов для выявления изменений в легких при COVID-19 по данным КТ

Наименование	Клиническая задача, решаемая ИИ-сервисом	Подготовительный этап (ретроспективное исследование) – признаки исследований калибровочного набора данных	Основной этап (проспективное исследование) – признаки, для которых ожидаются положительный и отрицательный результаты работы ИИ-сервиса	Содержание ответа ИИ-сервиса	Формат ответа ИИ-сервиса	Форма предоставления ответа ИИ-сервиса
Компьютерная томография органов грудной клетки	Выявление компьютерно-томографических признаков, коррелирующих с поражением легких при коронавирусной инфекции (COVID-19)	**Есть признаки патологии:** А*. 1. Инфильтрация легочной паренхимы по типу «матового стекла» с обеих сторон, преимущественно периферической локализации, с наличием или в отсутствие инфильтрации легочной паренхимы по типу консолидации с положительным признаком воздушной бронхограммы. 2. Инфильтрация легочной паренхимы по типу «булыжной мостовой» (утолщение междольковой интерстиция на фоне «матового стекла») с обеих сторон, преимущественно периферической локализации, с наличием или в отсутствие инфильтрации легочной паренхимы по типу консолидации с положительным признаком воздушной бронхограммы. Б. (исключительно для подготовительного этапа) 1. Положительные результаты лабораторной верификации коронавирусной инфекции (COVID-19) при помощи ПЦР-тестирования. 2. Установленный диагноз U07.1 (Коронавирусная инфекция, вызванная вирусом COVID-19, вирус идентифицирован). Для отнесения исследования к патологии достаточно одного из признаков. * На нативных изображениях		**Обязательно** – вероятность поражений легких, вызванных COVID-19 (признаки из списка А)	Число	Apache Kafka Message + DICOM SR
				Обязательно – классификация степени тяжести поражения легких по категориям «КТ 0–4» с указанием вероятности отнесения к каждому из классов	Класс поражения: КТ-0 – вероятность (число) КТ-1 – вероятность (число) КТ-2 – вероятность (число) КТ-3 – вероятность (число) КТ-4 – вероятность (число)	Apache Kafka Message + DICOM SR
				Обязательно – поражение (%) паренхимы отдельно для каждого лёгкого	Число	Apache Kafka Message + DICOM SR
		Нет признаков патологии: нет ни одного признака из перечня А		**Обязательно** – локализация найденных патологических находок	Контур/маска	DICOM

ИСТОЧНИКИ:

1. Лучевая диагностика коронавирусной болезни (COVID-19): организация, методология, интерпретация результатов: методические рекомендации / сост. С. П. Морозов, Д. Н. Проценко, С. В. Сметанина [и др.] // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 65. – М. : ГБУЗ «НПКЦ ДиТ ДЗМ», 2020. – 80 с. – URL: https://tele-med.ai/documents/500/19_ЛУЧЕВАЯ__ДИАГНОСТИКА__КОРОНАВИРУСНОЙ__БОЛЕЗНИ.pdf (дата обращения : 24.05.2021).
2. Министерство Здравоохранения Российской Федерации. Временные методические рекомендации. Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID-19). Версия 15 (22.02.2022)

Fundamental Principles of Standardization and Systematization of Information on Datasets

Innovative Technology for Structuring and Systematizing the Management of Medical Datasets for Instrumental Diagnostics:

- Terminology
- Registry Structure for Medical Instrumental Diagnostic Datasets
- Classification of Datasets according to the Purpose of Their Creation
- Classification of Data Verification Methods
- Principles of Naming

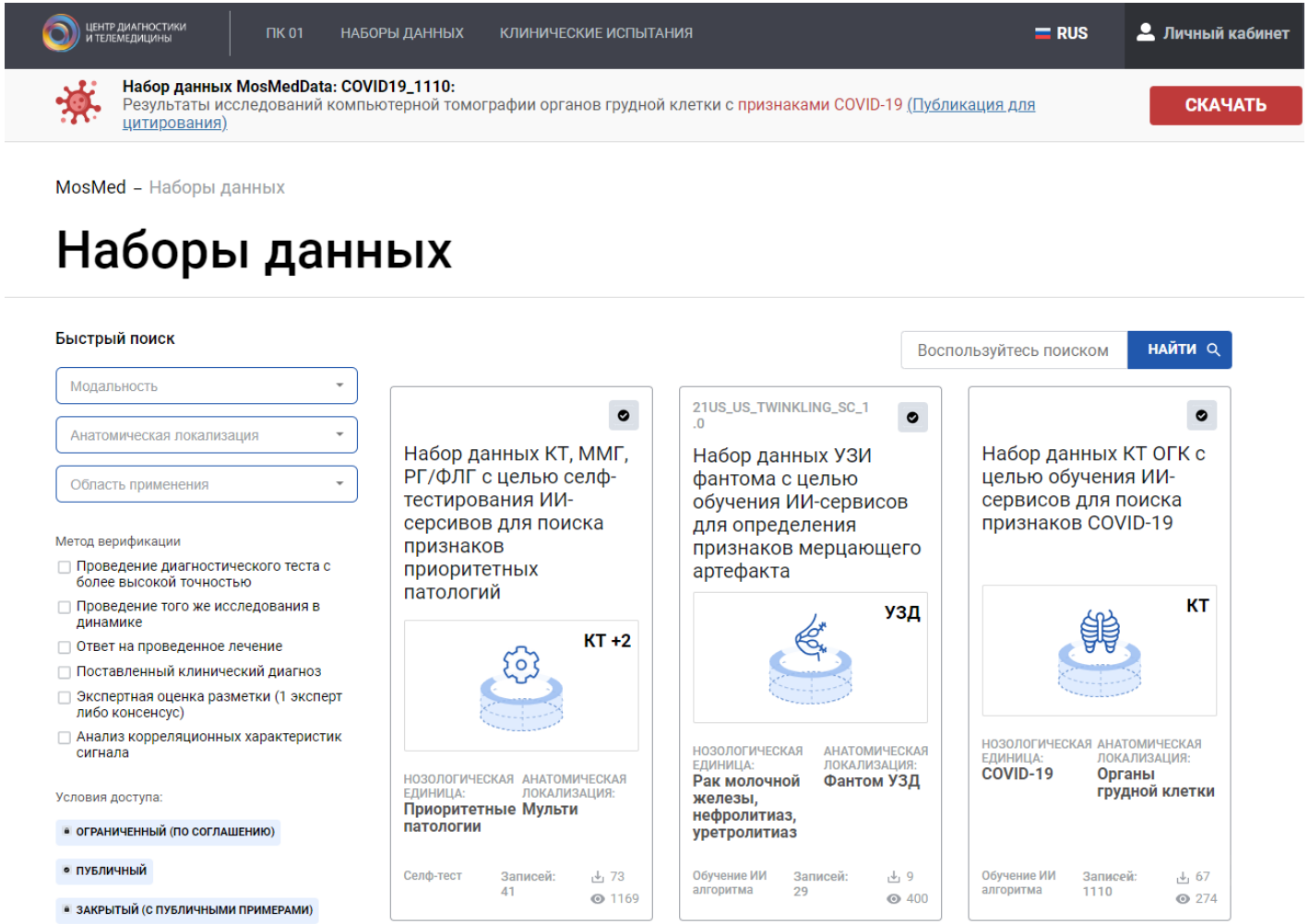
NB!

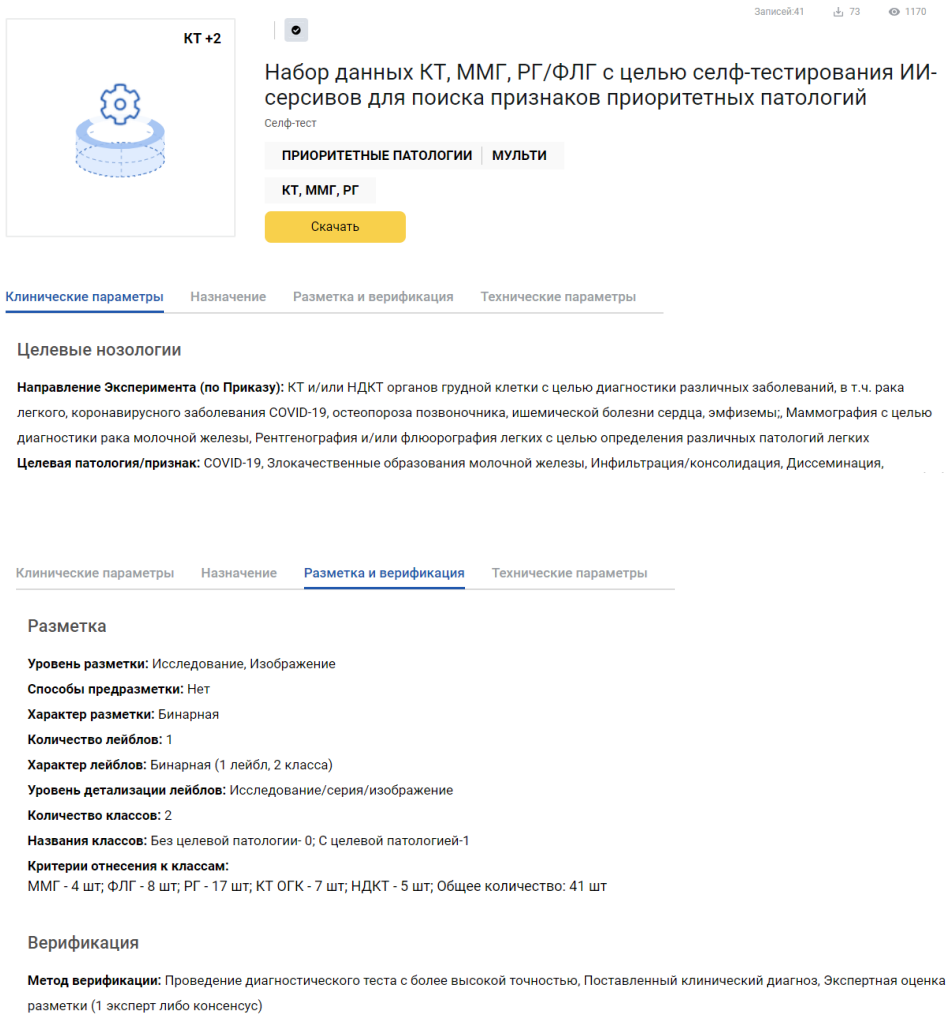
Approaches to sampling for quality control of ai in biomedical research

An analysis of global experience in regulations

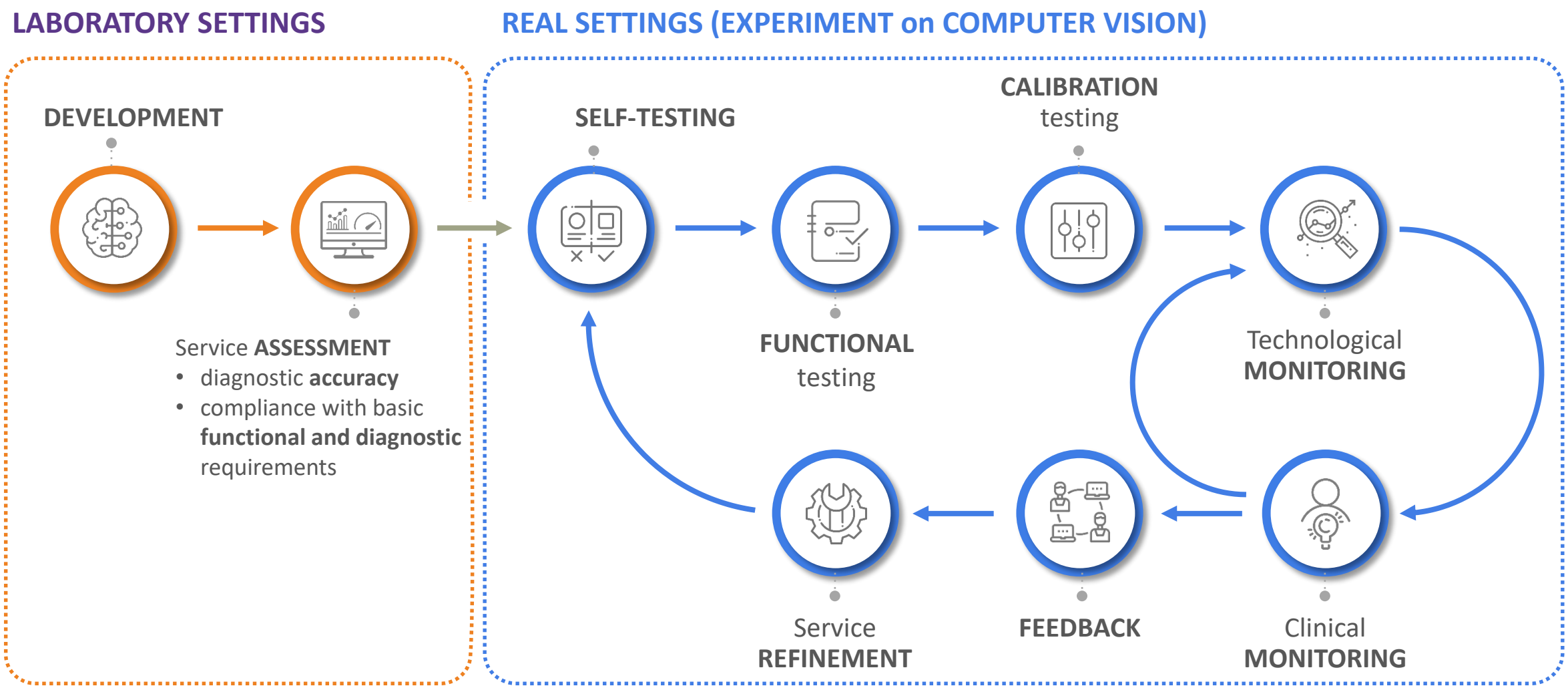
Vasilev et al, 2023, DOI: 10.21045/1811-0185-2023-4-28-41

Dataset library (mosmed.ai)





Life cycle of AI service in medical diagnostics



Quality control methods

TECHNICAL TRIALS

- **Compliance** with basic functional requirements
- Data exchange **integrity** between information systems
- Necessary supporting **information**
- **Presence and integrity** of the additional series (labeled images) and DICOM SR (draft reports)
- **Fundamental defects** (analysis of an incorrect anatomical area or projection, labelling outside the target organ, alteration of the original series)

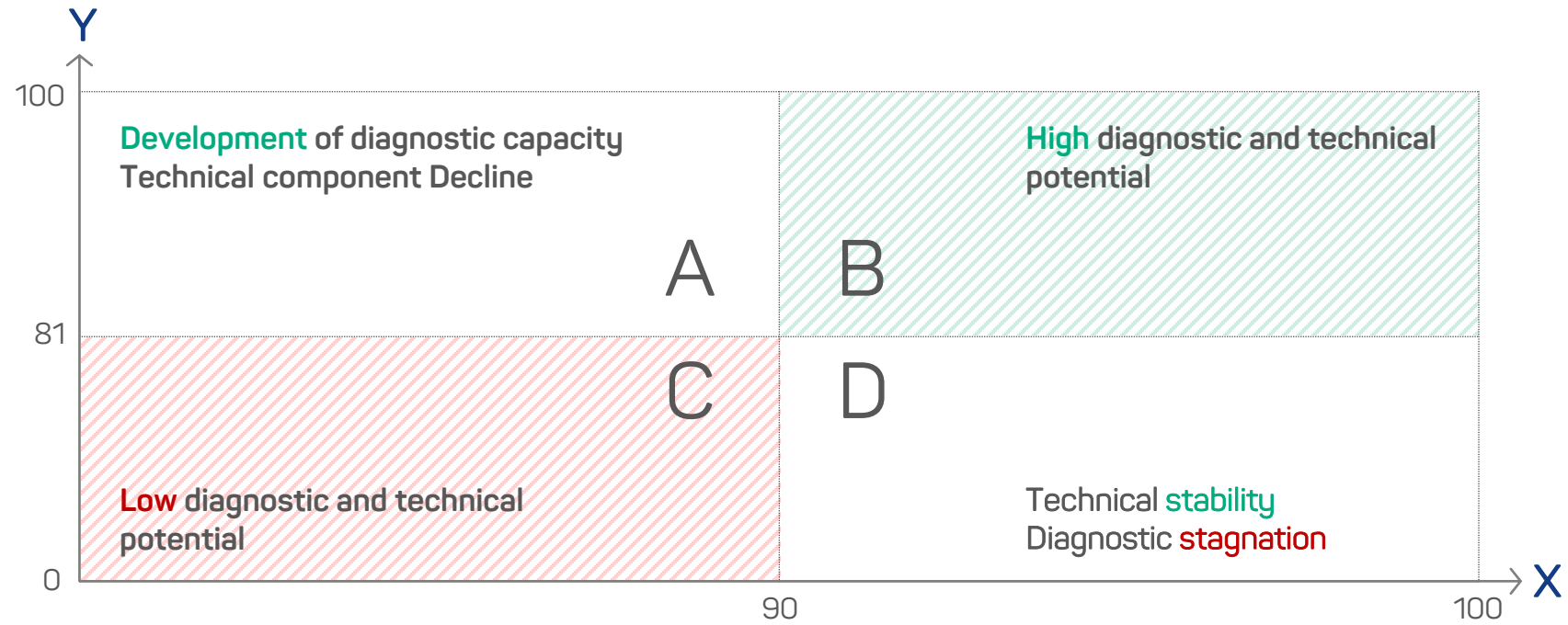
CLINICAL TRIALS

- **Compliance** with basic diagnostic requirements
- **Detection of target pathology** (binary solution)
- **Accuracy of labelling (outlining)**
- **Localization accuracy**
- **Accuracy of detection** of all pathological changes on the image

Qualitative component of the matrix

OY Axis is a clinical assessment (Prospective ROC AUC * 100)

Boundary "81" is a horizontal line with a boundary value for the clinical relevance of the "ROC AUC" in accordance with methodological recommendations (No. 43), equal to "0.81 * 100"



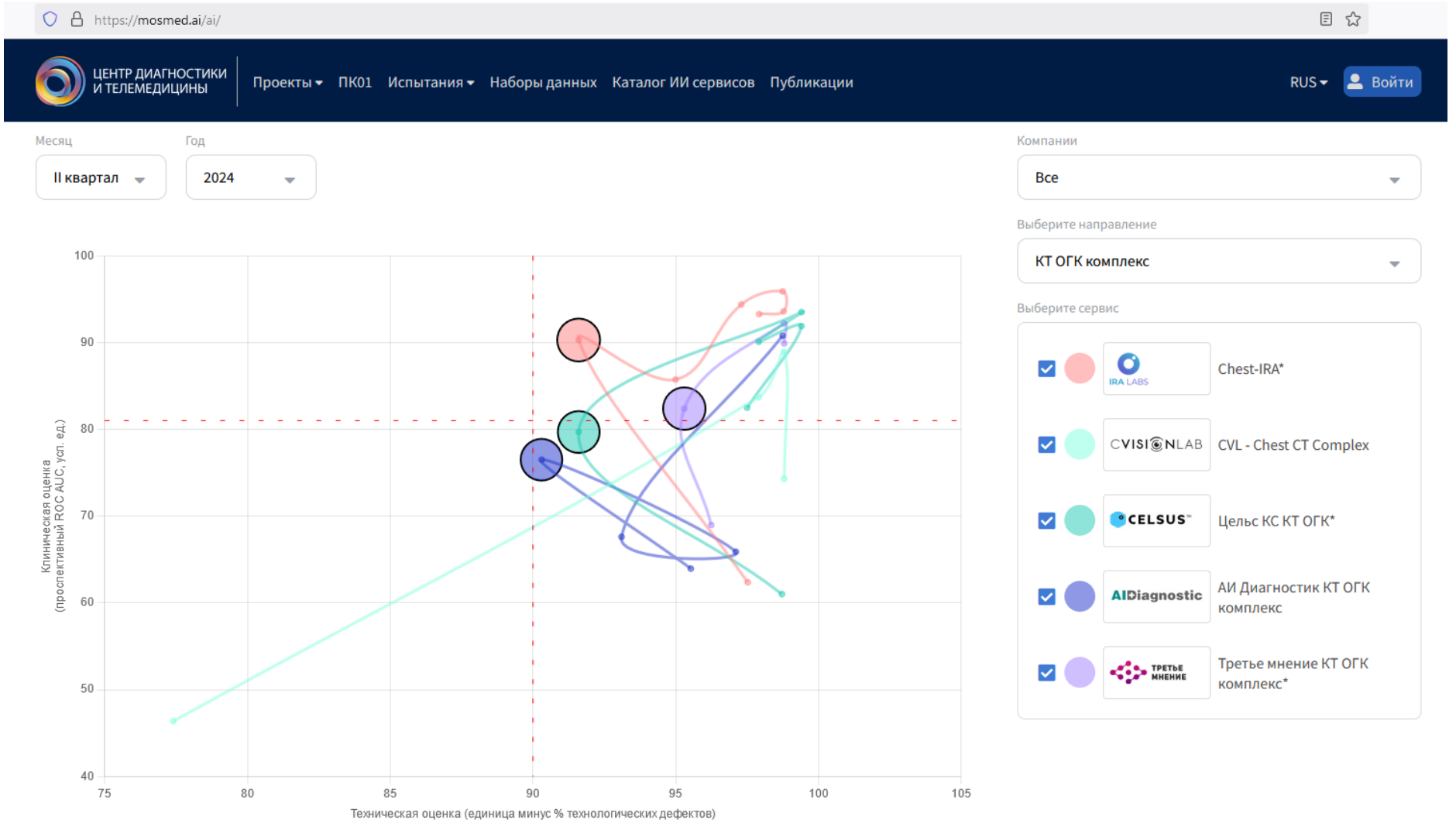
Increasing diagnostic accuracy

Increasing technical stability

OX Axis is a technical assessment (100 minus % technological defects)

Boundary "90" is a vertical line that reflects the mark of 10% technological defects, in accordance with the Order of the Moscow Healthcare Department

Maturity matrix: open information at MOSMED.AI



Rating methodology

As of 2023, a revised approach to assessing AI services has been implemented. This new methodology presents qualitative performance indicators for AI services within a maturity matrix, categorized by modalities. The maturity matrix was developed and finalized by the end of 2022, with provisions for quarterly updates.

Moscow Experiment on the Application of Computer Vision Technologies in Diagnostic Radiology

World's largest multicenter prospective clinical trial of AI in radiology

> 16 m
Imaging studies

> 40
modalities

> 150
Healthcare Facilities

~ 1 800
radiologists

> 30
AI developers

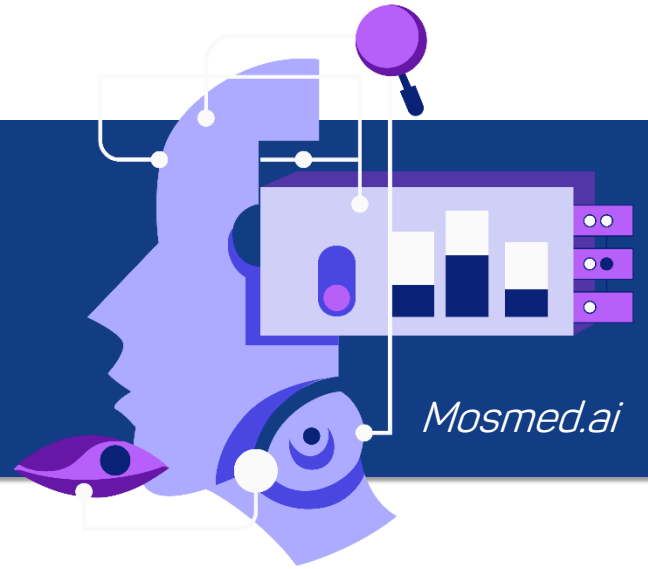
~ 380
datasets

> 50
AI services **now**

> 200
AI services **in 5 years**

2023
Special rates for AI analysis of X-ray studies have been established in the Regional Compulsory Health Insurance

2024
This Experiment has been EXPANDED to the regions of Russia



DICOM SR – text description of the study

2020

Назначение сервиса

Модальность: ФЛГ, РГ
Анатомическая область: грудная клетка
Пациенты: взрослое население
Назначение: поиск признаков патологий в легких
Проекция снимка: прямая

Краткое руководство пользователя

[redacted]автоматическая система, выявляющая области на грудной клетки, в которых предполагается наличие патологий. Области локализуются красными контурами на снимке. Вероятность наличия патологии характеризуется численным значением в подписи к снимку.

Заключение Вероятность патологических изменений - 0.78

Краткое руководство пользователя

Выявленные изменения органов грудной клетки отмечены прямоугольником с контуром, ограничивающим соответствующий выделенной области фрагмент горячей карты.
Вероятность наличия выявленного изменения находится в подписи к каждой находке.
Сервис не уведомляет, если исследование проведено технически неверно.

Заключение Вероятность наличия патологического изменения органов грудной клетки в данном исследовании 42%
«Находка 1» с вероятностью 76.89 Percent
«Находка 1»

2024

Заключение:

Выявлены патологические изменения. Вероятность 100%.
Обнаружены следующие рентгенографические признаки:
Изменения плевры

Заключение

Вероятность наличия патологических изменений: 0.79
Выявленные патологические области:
1. Класс патологии не определен
2. Инфильтрация/консолидация
3. Плевральный выпот
Количество выявленных патологических изменений: 3

Описание

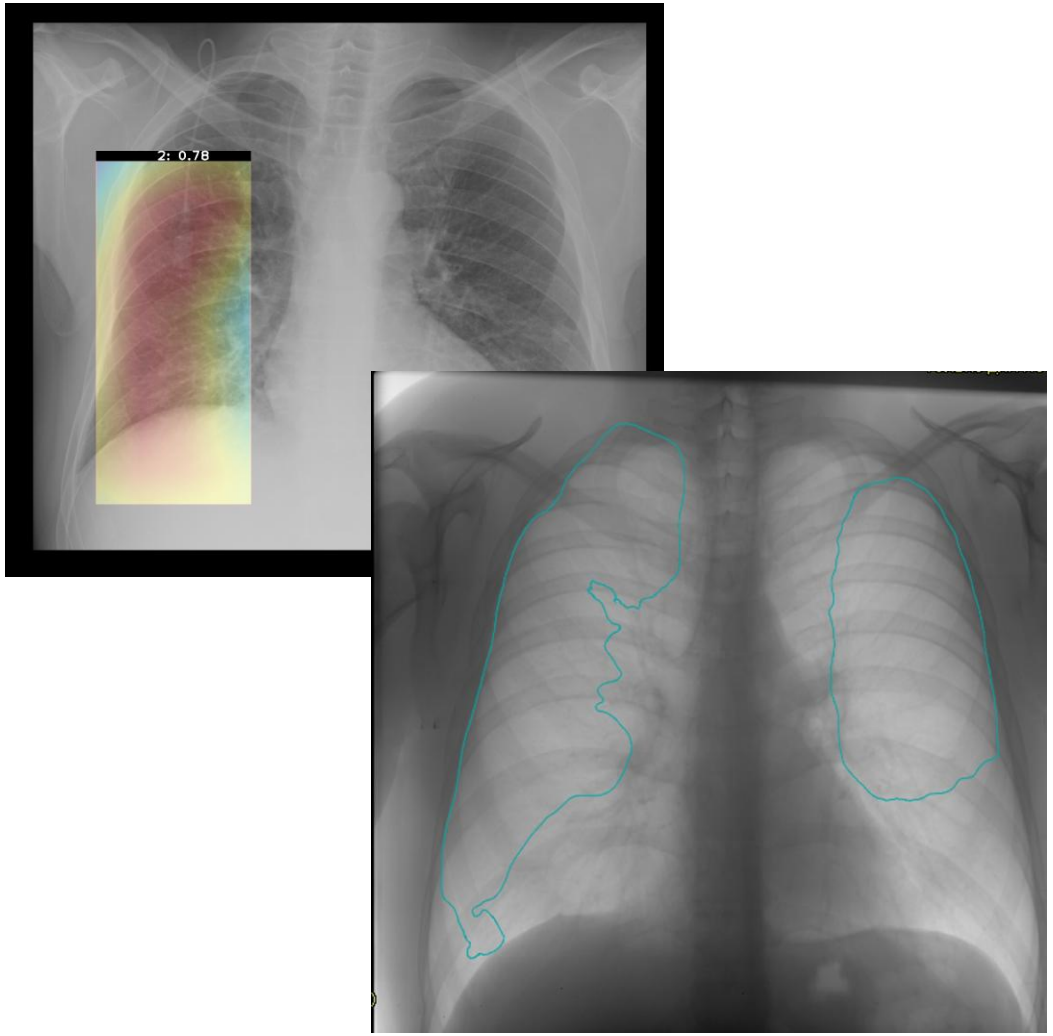
Находка EFN, Плевральный выпот (зеленый контур) с вероятностью 0.12; МКБ: A15, A16, C34-39, J10-18, J80-84, J94, J98, Z00, Z12, R09
Находка PTX, Пневмоторакс (красный контур) с вероятностью 0.93; МКБ: S27, Z00, Z12
Находка ATS, Ателектаз (желтый контур) с вероятностью 0.34; МКБ: A15, A16, C34-39, Z00, Z12, R09
Находка FCS, Очаг затемнения (розовый контур) с вероятностью 0.32; МКБ: A15, A16, C34-39, J10-18, J80-84, J94, J98, Z00, Z12, R09
Находка CAV, Полость с распадом или с уровнем жидкости (оранжевый контур) с вероятностью 0.28; МКБ: A15, A16, C34-39, J10-18, Z00, Z12, R09
Находка CSN, Инфильтрация/консолидация (коричневый контур) с вероятностью 0.91; МКБ: A15, A16, C34-39, J10-18, J80-84, J94, J98, Z00, Z12, R09
Находка DSM, Диссеминация (фиолетовый прямоугольник) с вероятностью 0.48; МКБ: A15, A16, C34-39, J10-18, Z00, Z12, R09

Заключение Вероятность наличия патологических изменений органов грудной клетки в данном исследовании 0.86

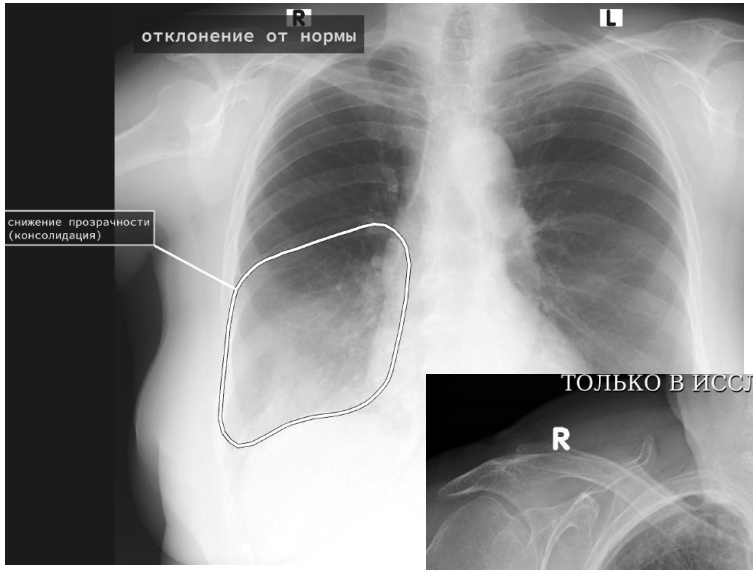
DICOM SC – additional series with contouring pathological changes

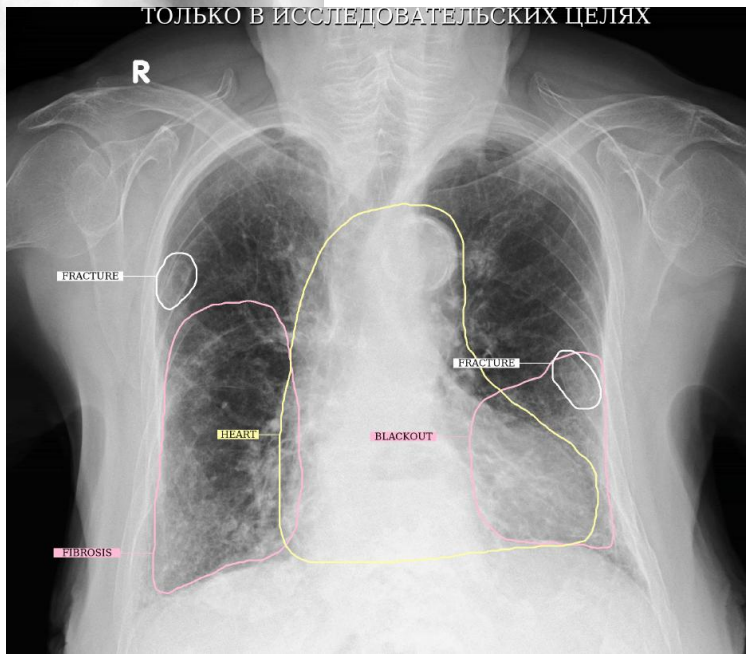
telemedai.ru

2020

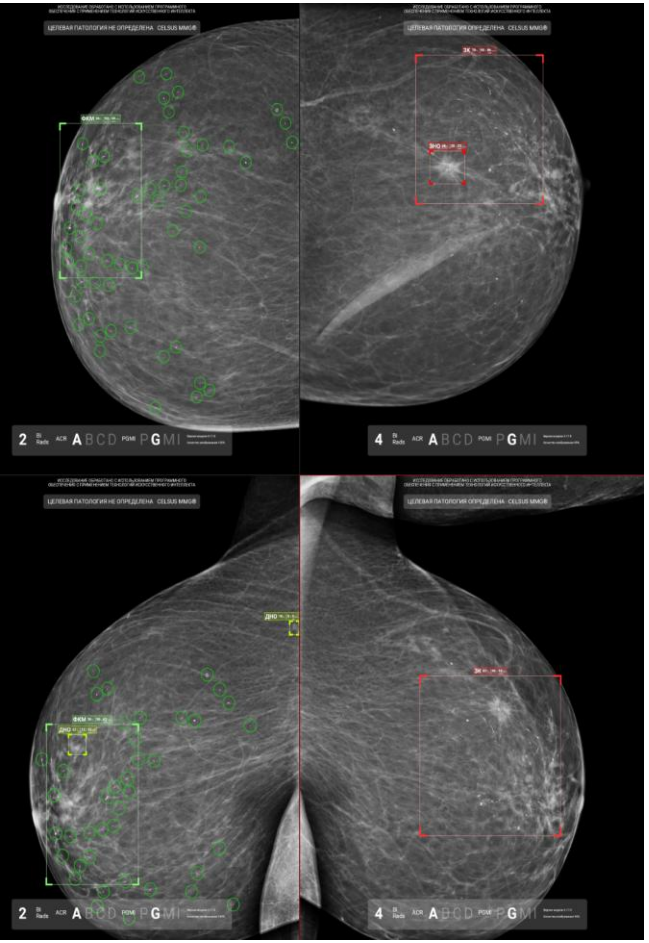


2024





AI Results



Описание:

Область целевой патологии выделена красным боксом с лейблом и аббревиатурами ЗНО и ЗК.
Вероятность патологии - 0,95

Сторона L: Структура молочной железы по ACR - A. Молочная железа практически полностью жировой плотности. Маммография обладает высокой чувствительностью

Вероятность наличия злокачественных изменений в молочной железе - 0,95

Проекция: CC
Определены:
Злокачественный кальцинат, вероятность находки - 0,70, вероятность злокачественности - 0,70, площадь 7740 кв. мм, высота = 90 мм, ширина = 86 мм
Злокачественное новообразование, вероятность находки - 0,68, вероятность злокачественности - 0,68, площадь 460 кв. мм, высота = 20 мм, ширина = 23 мм

Проекция: MLO
Определены:
Злокачественный кальцинат, вероятность находки - 0,67, вероятность злокачественности - 0,67, площадь 9120 кв. мм, высота = 96 мм, ширина = 95 мм

Сторона R: Структура молочной железы по ACR - A. Молочная железа практически полностью жировой плотности. Маммография обладает высокой чувствительностью

Вероятность наличия злокачественных изменений в молочной железе - 0,34

Проекция: CC
Определены:
Доброкачественное новообразование (фиброзно-кистозная мастопатия), вероятность находки - 0,26, площадь 5115 кв. мм, высота = 93 мм, ширина = 55 мм

Проекция: MLO
Определены:
Доброкачественное новообразование, вероятность находки - 0,18, площадь 48 кв. мм, высота = 8 мм, ширина = 6 мм
Доброкачественное новообразование (фиброзно-кистозная мастопатия), вероятность находки - 0,36, площадь 5952 кв. мм, высота = 96 мм, ширина = 62 мм
Доброкачественное новообразование, вероятность находки - 0,22, площадь 144 кв. мм, высота = 12 мм, ширина = 12 мм

Оценка категории по диагностической шкале BI-RADS: для правой молочной железы BI-RADS 2, для левой молочной железы BI-RADS 4
Скрининг оценка категории: для правой молочной железы BI-RADS 2, для левой молочной железы BI-RADS 0

Заключение:

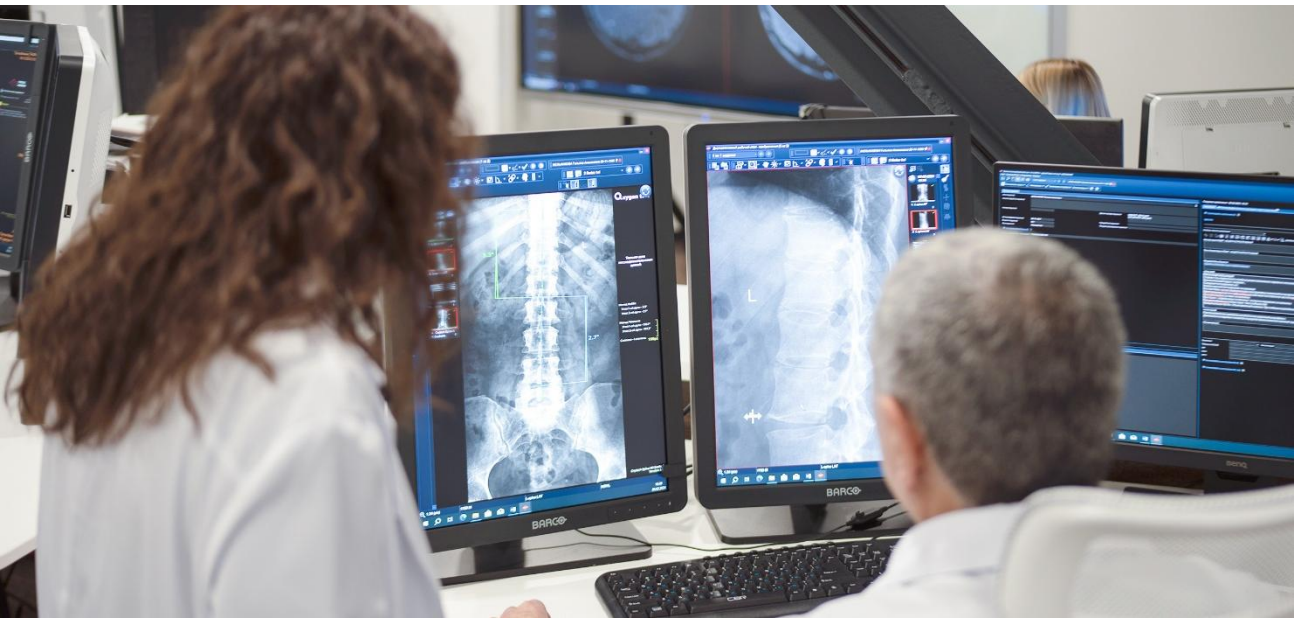
Вероятность патологии - 0,95.
Оценка категории по диагностической шкале BI-RADS: для правой молочной железы BI-RADS 2, для левой молочной железы BI-RADS 4
Скрининг оценка категории: для правой молочной железы BI-RADS 2, для левой молочной железы BI-RADS 0

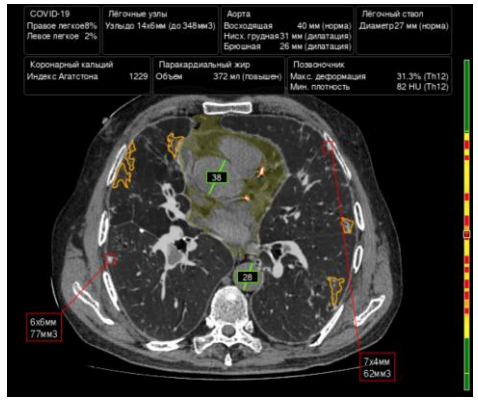
Сторона L: Тип плотности по ACR - A.
Выявлены признаки злокачественных изменений в левой молочной железе

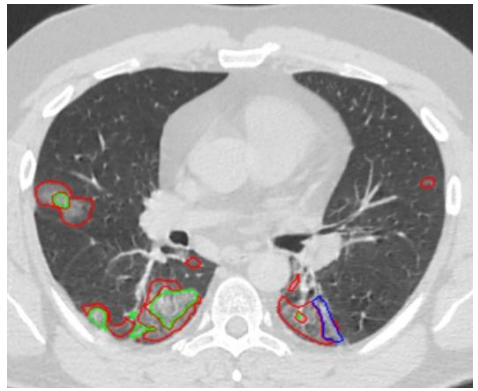
Сторона R: Тип плотности по ACR - A.
Признаки злокачественных изменений в правой молочной железе не выявлены

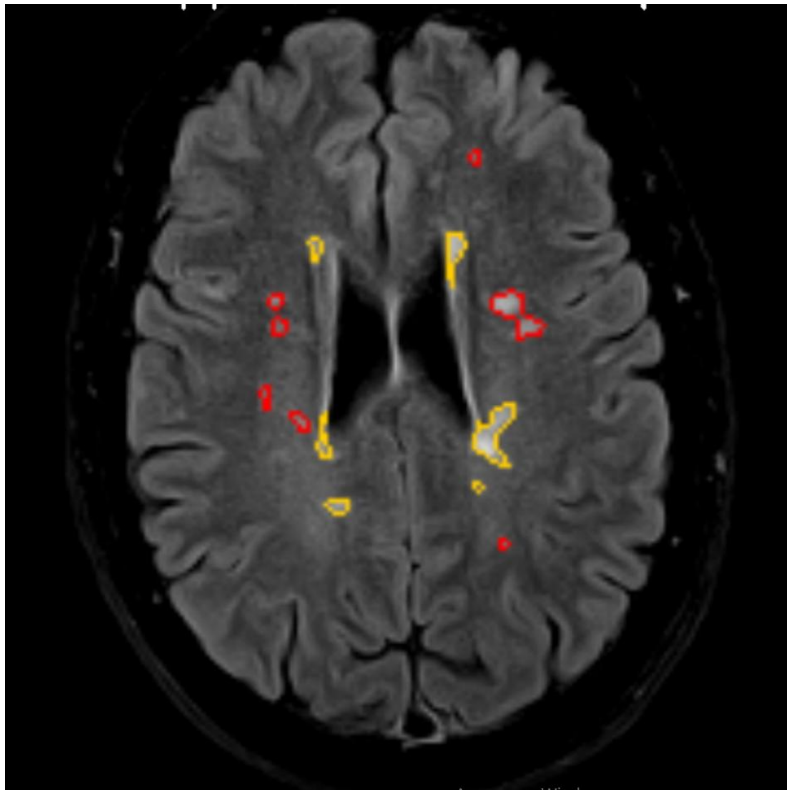
- Original Images
- AI-labeled Images
- AI Short Report (DICOM SR)

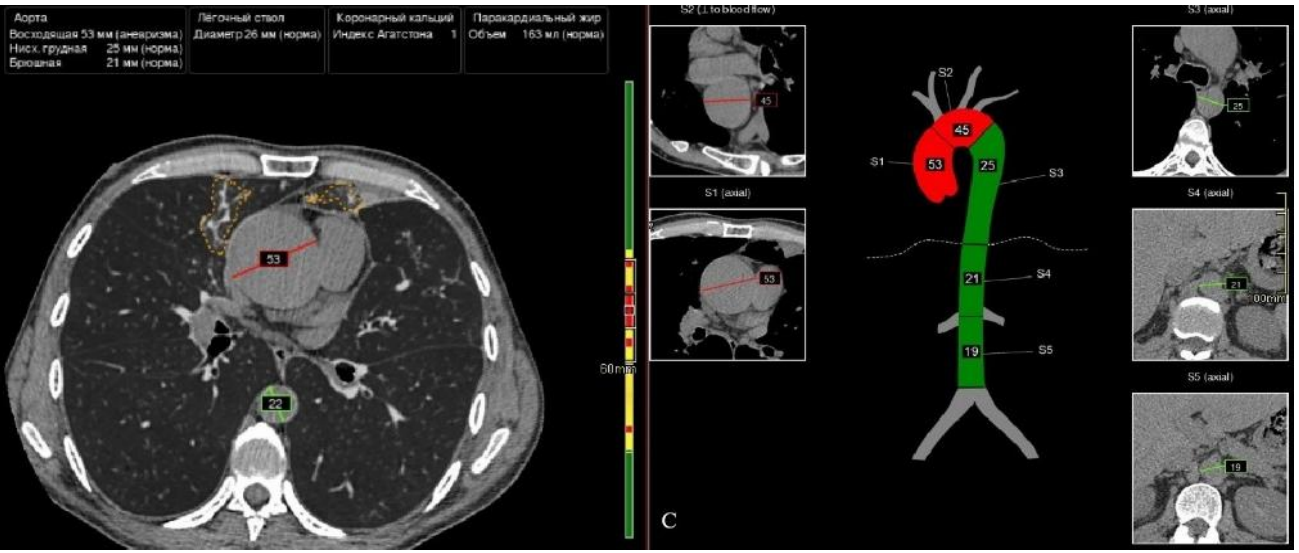
Routine use of AI



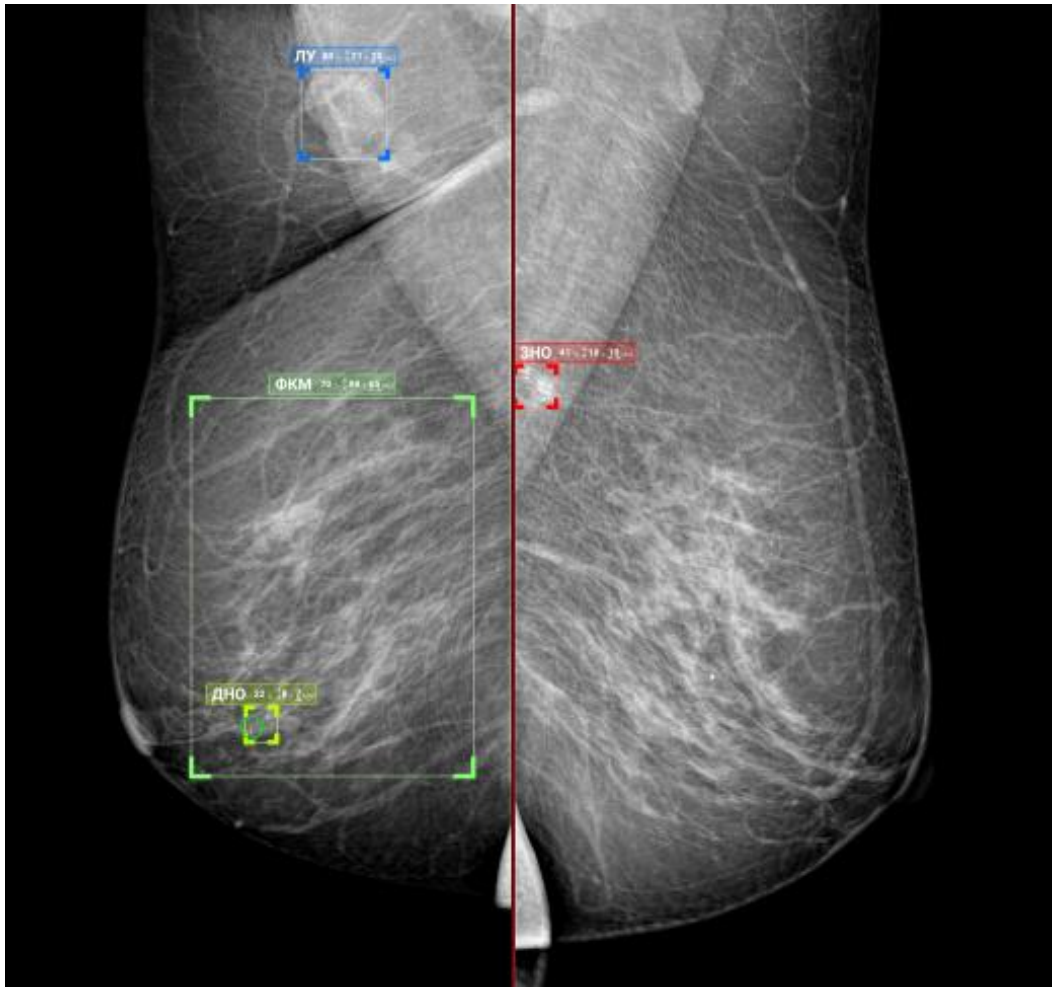








AI-Assisted Double Reading Covered by Universal Health Coverage



In 2023, Moscow became the first Russian region to implement AI-assisted breast cancer screening with radiologist review (as medical service in Universal Health Coverage)

15%
reduction in report turnaround time

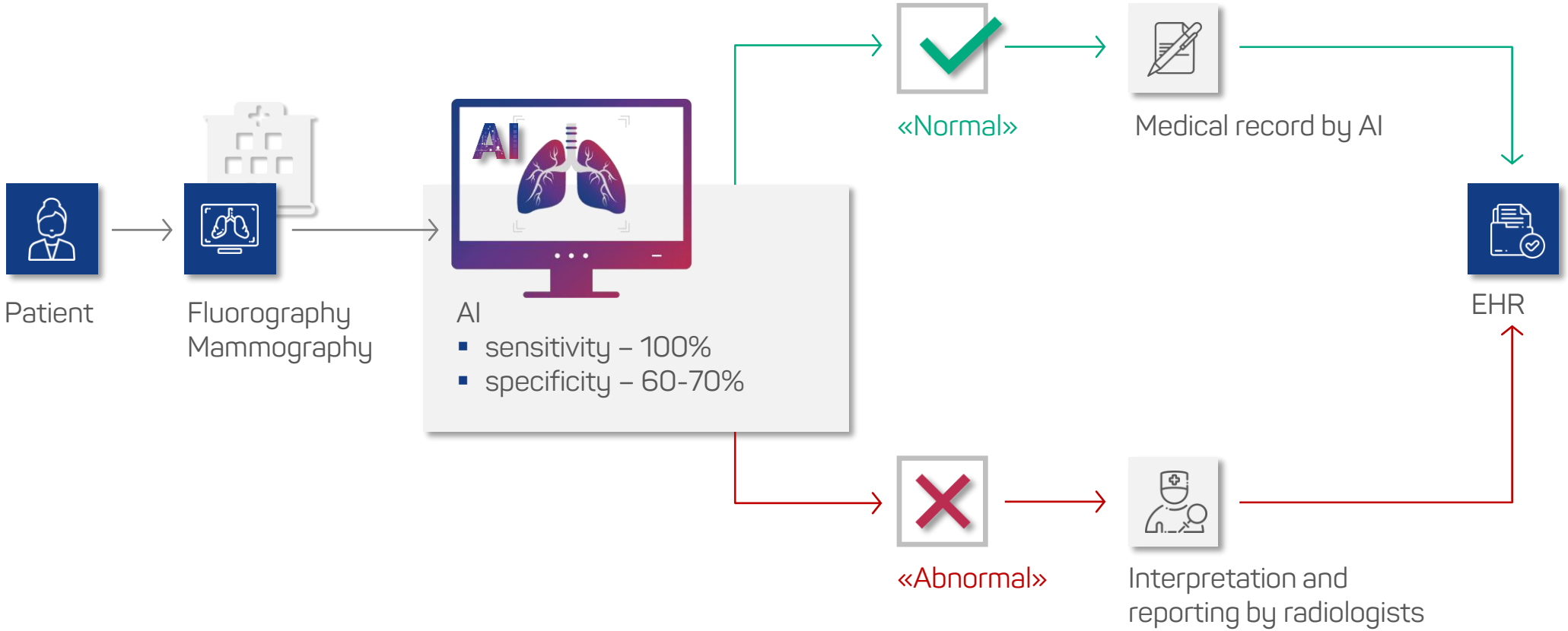
88%
decrease in overall production process duration

654 731
AI+Radiologist co-readings performed in 2023-2024

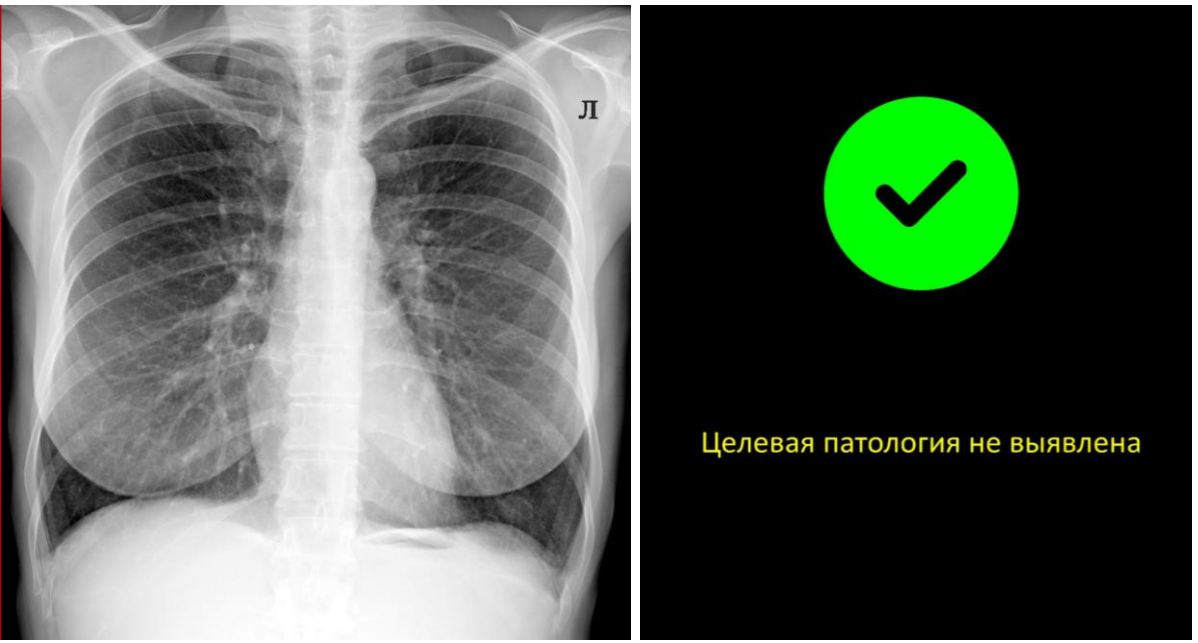
32%
reduction in operational costs

AI-assisted breast cancer **screening demonstrated** a **0,1%** underdiagnosis rate

Autonomous Triage of Screening Chest X-ray



Impact of Autonomous CXR Results Triage



World's First Implementation of Autonomous Triage for Mass Preventive Screening Results

575 549 CXR studies underwent **autonomous triage** in 2024

54,8% **reduction in radiologist workload** (no human interpretation required)

Triage accuracy reached 99,95%

Financial costs were **reduced** by **43,7%**

Vasilev YuA, Sychev DA, Bazhin AV, Shulkin IM, Vladzymyrskyy AV, Golikova AYu, Arzamasov KM, Mishchenko AV, Bekdzhanyan GA, Goldberg AS, Rodionova LG. Autonomous artificial intelligence for sorting results of preventive radiological examinations of chest organs: medical and economic efficiency. Digital Diagnostics. 2025;6(1):5–22.
DOI: https://doi.org/10.17816/DD641703.

Subcommittee 01/TC 164 ‘AI in Healthcare’: Structure

TC 164
‘ARTIFICIAL INTELLIGENCE’

SC 01
‘AI IN HEALTHCARE’
Center for Diagnostics & Telemedicine

44
MEMBERS

- program for the development of national standards
- introduction of international and regional standards
- scientific, technical, legal and regulatory assessment of draft standards
- endorsement and establishment of task teams within the Subcommittee S 42 ‘Artificial Intelligence’ (ISO/IEC JTC 1 ‘Information Technologies’)
- projects and proposals for the development of international standards
- issuing conclusions on feasibility of international, regional, national standards and codes of practice of foreign states

Key members of the Subcommittee:

24
agencies and companies

+ 20
external experts

GOST R national standards under way in 2025

General requirements for AI in medicine

Big data

AI Clinical Trials and MD Certification

AI systems in radiology

Analytics and forecasting

Clinical Decision Support System

Telemonitoring

2024 | **TOTAL 15** GOST R standards | **10** final versions **+ 5** pilot versions

2025 | **TOTAL 6** GOST R standards | **6** final versions

Recognizing and Progress of the Moscow Experiment: MosMed.AI

President of Russian Federation V.V. Putin visited the Center for Diagnostics and Telemedicine

State-wide Scaling: MosMedAI

2024

72
regions are connected to the platform

- organizational and methodological support, alongside quality assurance measures
- digital infrastructure modernization for radiology workflows
- outcome validation required

~5 m
studies processed

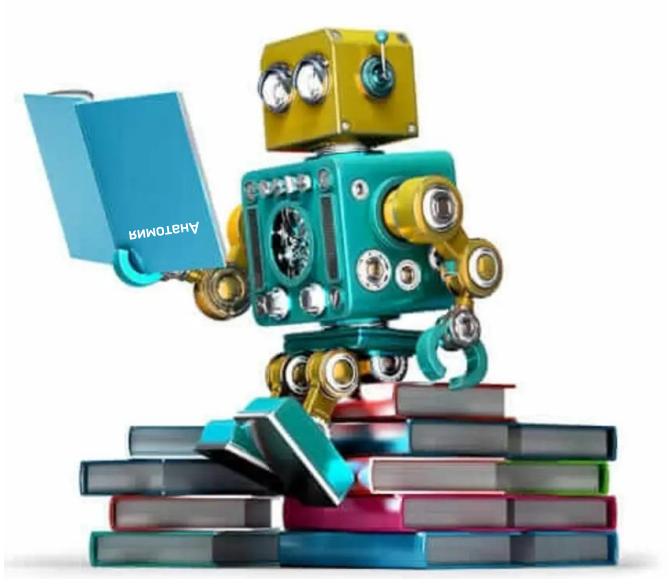
~1 300
health facilities

5
types of studies

GPTing or not GPTing

LLM/GPT successfully passed the medical exams

- ✓ for admission to medical practice
- ✓ obstetrics and gynecology
- ✓ orthopedics
- ✓ urology
- ✓ general surgery
- ✓ otolaryngology
- ✓ pharmacy
- ✓ dentistry
- ✓ psychiatry
- ✓ nursing (haha)



Average correct answers to **70%** of questions

Jin HK, Lee HE, Kim E. Performance of ChatGPT-3.5 and GPT-4 in national licensing examinations for medicine, pharmacy, dentistry, and nursing: a systematic review and meta-analysis. BMC Med Educ. 2024 Sep 16;24(1):1013. doi: **10.1186/s12909-024-05944-8.**
Liu M, Okuhara T, Chang X, Shirabe R, Nishiie Y, Okada H, Kiuchi T. Performance of ChatGPT Across Different Versions in Medical Licensing Examinations Worldwide: Systematic Review and Meta-Analysis. J Med Internet Res. 2024 Jul 25;26:e60807. doi: **10.2196/60807**.

The same people who criticize the Unified State Exam talk about a success of GPT

Well, what does it mean, "passed"? ... It picked up the answers from textbooks.

And what about a patient examination, practical skills, and clinical data?

Purpose and objectives of the study

PURPOSE

Substantiation and development of a prototype system for automating the basic radiologist's workflow (AI assistant) based on the large-scale generative models

OBJECTIVES

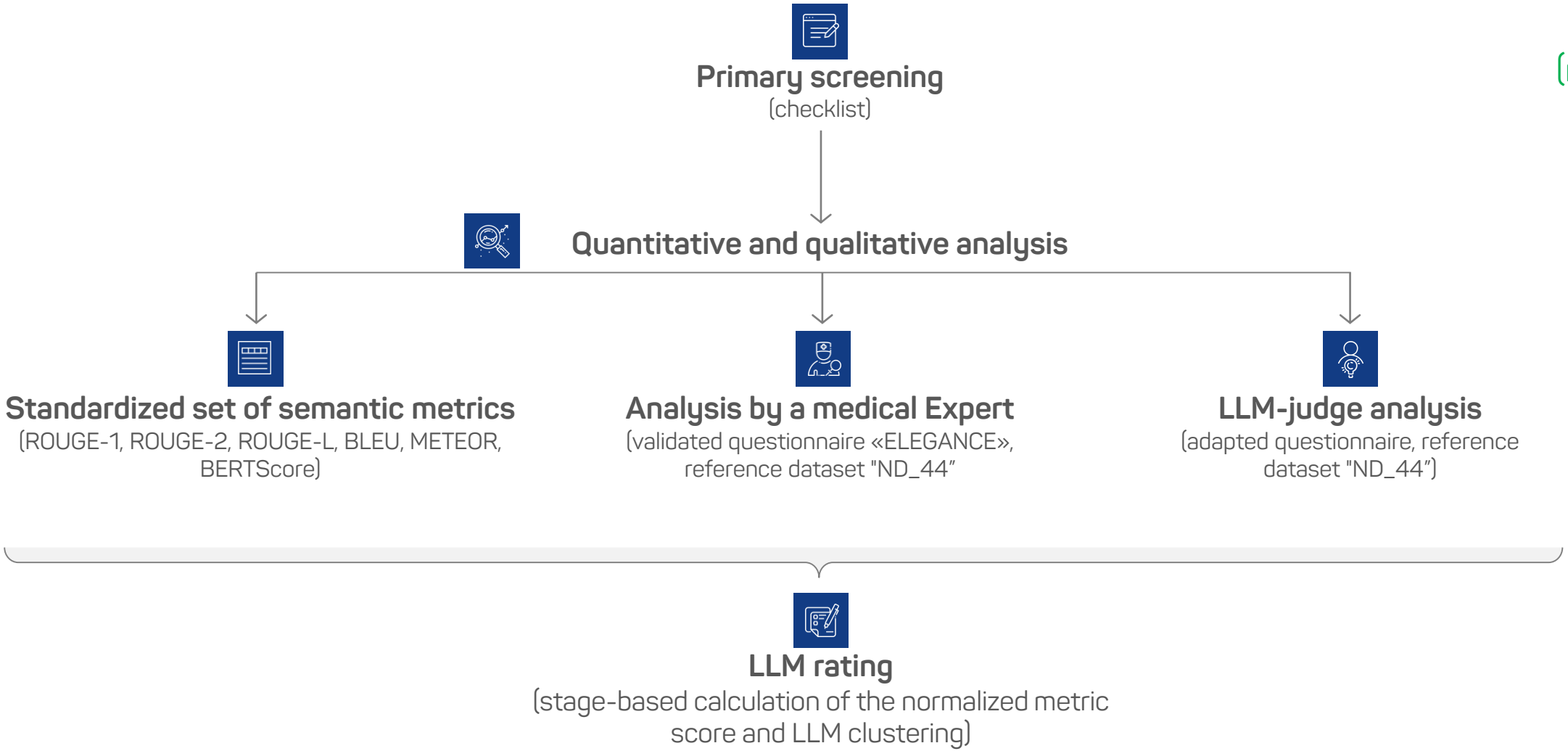
1. **To analyze existing approaches** to the LLM application in diagnostic radiology, including the assessment of their capabilities and limitations
2. **To substantiate functions of the AI assistant** to a radiologist, taking into account the clinical context in outpatient and inpatient settings
3. **To develop a methodology and tools for assessing the LLM quality** applied in healthcare
4. **To conduct LLM testing and selection** using the developed methodology
5. **To develop a methodology and tools for creating and verifying** datasets for testing and assessing LLM quality in healthcare
6. **To substantiate the optimal architecture** of the AI assistant to a radiologist by testing and selecting LLM using the developed methodologies
7. **To develop a software prototype,** conduct pilot testing, and evaluate its effectiveness

Author's set of metrics for assessing and selecting LLMs

	Formal criteria	Technical (semantic) metrics	Expert assessment by radiologists	LLM judge	Assessment summary
Tool	Checklist*	▪ ROUGE-1, ROUGE-2, ROUGE-L ▪ BLEU, METEOR, BERTscore ▪ Scripts for calculation*	ELEGANCE questionnaire (Expert-Led Evaluation of Generative AI Competence & Excellence) *	▪ Justified model selection* ▪ Questionnaire for automated control*	▪ Normalized metric score (NMS)*
Composition	▪ Architecture type ▪ Open access ▪ Size ▪ Relevance ▪ Adaptation to the Russian language ▪ Ability to work with a large context	▪ Handbook of calculation methods ▪ Dashboard*	Validated domain questionnaire*: ▪ relevance ▪ extensiveness ▪ appropriateness ▪ coherence and structure ▪ grammar and terminology ▪ hallucinations ▪ completeness ▪ redundancy	▪ Questionnaire composition* – relevance – completeness – redundancy – coherence and structure – grammar and terminology – hallucinations	Clustering

** Original scientific development of the Center for Diagnostic and Telemedicine*

Methodology of comparative study and selection of LLM



LLM clusters

LLM judge

Cluster 1 "Best"	Cluster 2 "Average"	Cluster 3 "Worst"
Qwen3-14B-nt	Qwen2.5-32B-Instruct_Q4	RuadaptQwen2.5-32B-Pro-Beta_Q2
Qwen3-14B_Q4-nt	T-pro-1.0	cogito 32b_Q4
THUDM_GLM-4-32B_Q4	Qwen3-4B	Mistral-Nemo-Instruct-2407
Qwen3-32B_Q4	gemma-3-27b-it_Q4	saiga_nemo_12b
Qwen3-30B-A3B_Q4	RuadaptQwen2.5-32B-Pro-Beta_Q8	gemma-2-27b-it
Qwen3-32B_Q4-nt	gemma-3-12b-it	gemma-2-9b-it
Qwen3-30B-A3B_Q4-nt	Vikhr-Nemo-12B-Instruct-R-21-09-24_Q4	DeepSeek-R1-Distill-Llama-Q4
Qwen3-14B	gemma-3-12b-it_Q4	Qwen2-7B-it
YandexGPT-5-Lite-8B-instruct_Q4	llama3.1-8b	Phi-4-mini-instruct
Mistral-Small-3.1-24B-Instruct-2503_Q4	Vikhr-YandexGPT-5-Lite-8B	granite-3.2-8b-instruct
Qwen2.5-72B-Instruct_Q4	GigaChat-20B-A3B-instruct-v1.5_Q4	deepseek-v2_16b_Q8
Qwen3-8B	Qwen2.5-7B-Instruct	Llama-3.2-3B-Instruct
Llama-3.3-70B-Instruct_Q4	DeepSeek-32B	granite-3.3-8b_Q4
Qwen3-14B_Q4	phi-4	
QwQ-32B_Q4	cogito 14b_Q4	
Qwen2.5-14B-Instruct		

Base (core)

Cluster 1 "Best"	Cluster 2 "Average"	Cluster 3 "Worst"
Qwen3-14B-nt	QwQ-32B_Q4	Qwen2-7B-it
Qwen3-14B_Q4-nt	Qwen2.5-14B-Instruct	Phi-4-mini-instruct
THUDM_GLM-4-32B_Q4	Qwen2.5-32B-Instruct_Q4	granite-3.2-8b-instruct
Qwen3-32B_Q4	T-pro-1.0	deepseek-v2_16b_Q8
Qwen3-30B-A3B_Q4	Qwen3-4B	Llama-3.2-3B-Instruct
Qwen3-32B_Q4-nt	gemma-3-27b-it_Q4	granite-3.3-8b_Q4
Qwen3-30B-A3B_Q4-nt	RuadaptQwen2.5-32B-Pro-Beta_Q8	
Qwen3-14B	gemma-3-12b-it	
YandexGPT-5-Lite-8B-instruct_Q4	llama3.1-8b	
Mistral-Small-3.1-24B-Instruct-2503_Q4	Vikhr-YandexGPT-5-Lite-8B	
Qwen2.5-72B-Instruct_Q4	GigaChat-20B-A3B-instruct-v1.5_Q4	
Qwen3-8B	Qwen2.5-7B-Instruct	
Llama-3.3-70B-Instruct_Q4	DeepSeek-32B	
Qwen3-14B_Q4	phi-4	
	cogito 14b_Q4	
	RuadaptQwen2.5-32B-Pro-Beta_Q2	
	cogito 32b_Q4	
	Mistral-Nemo-Instruct-2407	
	saiga_nemo_12b	
	gemma-2-27b-it	
	gemma-2-9b-it	
	DeepSeek-R1-Distill-Llama-8B_Q4	

LLM name	Category	Number of parameters, billion pcs.	Context window, thousand tokens	License
Mistral-small-24b	LLM judge	24	128	Apache 2.0
Qwen3-14b	Core	14	128	Apache 2.0
Qwen3-32b-Q4	Core	32	128	Apache 2.0

Section on LLM in Medicine and
AI Assistant Development
October 9–10, 2025
itmcongress.ru

CENTER FOR DIAGNOSTICS AND TELEMEDICINE

MOSCOW HEALTHCARE

THANK YOU FOR YOUR ATTENTION

npcmr@zdrav.mos.ru +7 (495) 276 - 04 - 36 telemedai.ru