

# ОТ ВНИМАНИЯ ВРАЧА К ДИАГНОСТИКЕ: ИНТЕГРАЦИЯ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ ДЛЯ ПРОГНОЗИРОВАНИЯ ВЗГЛЯДА ВРАЧА

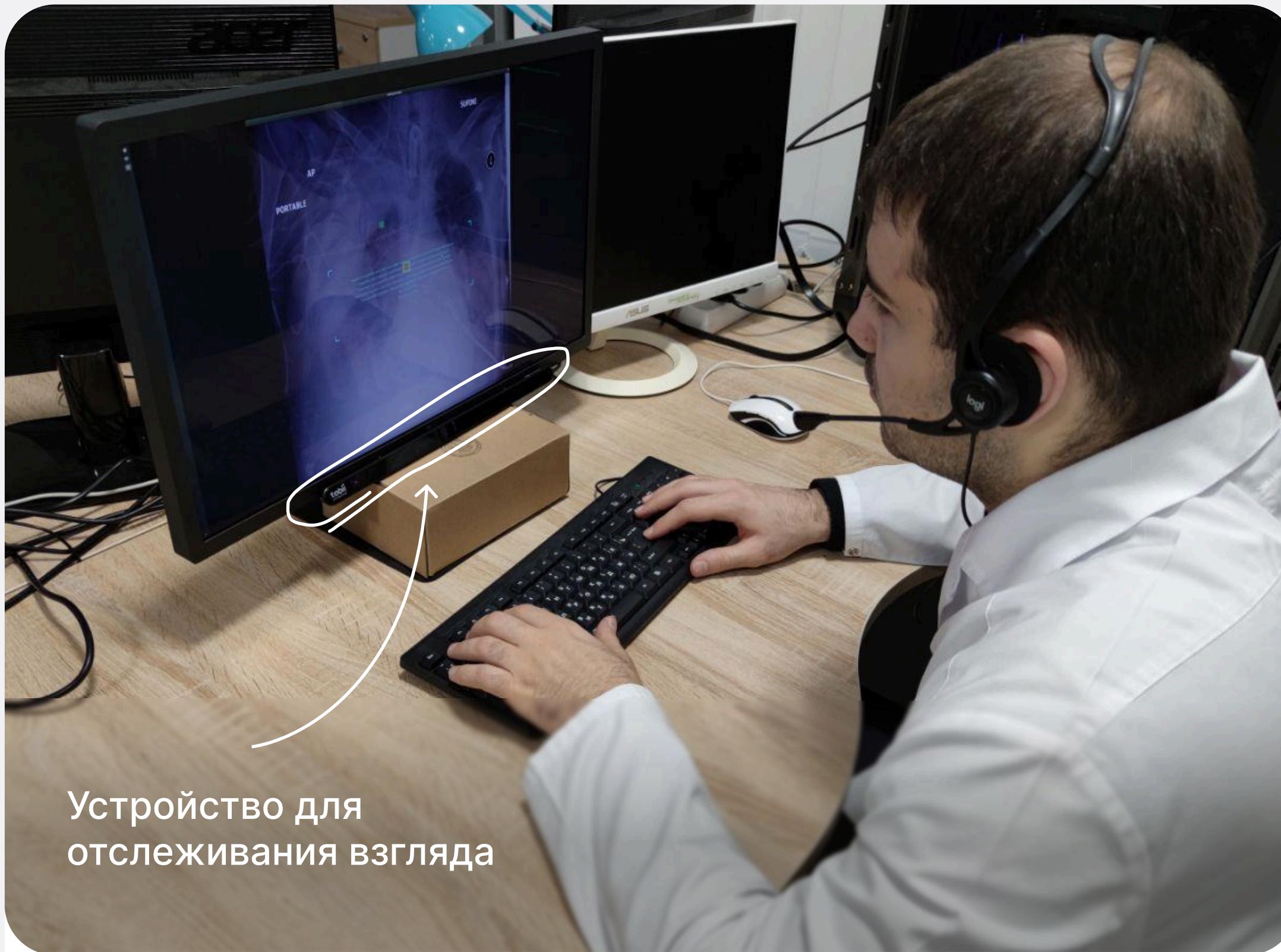
Першин Илья (докладчик)  
Львов Дмитрий

Лаборатория ИИ в медицине  
Университет Иннополис

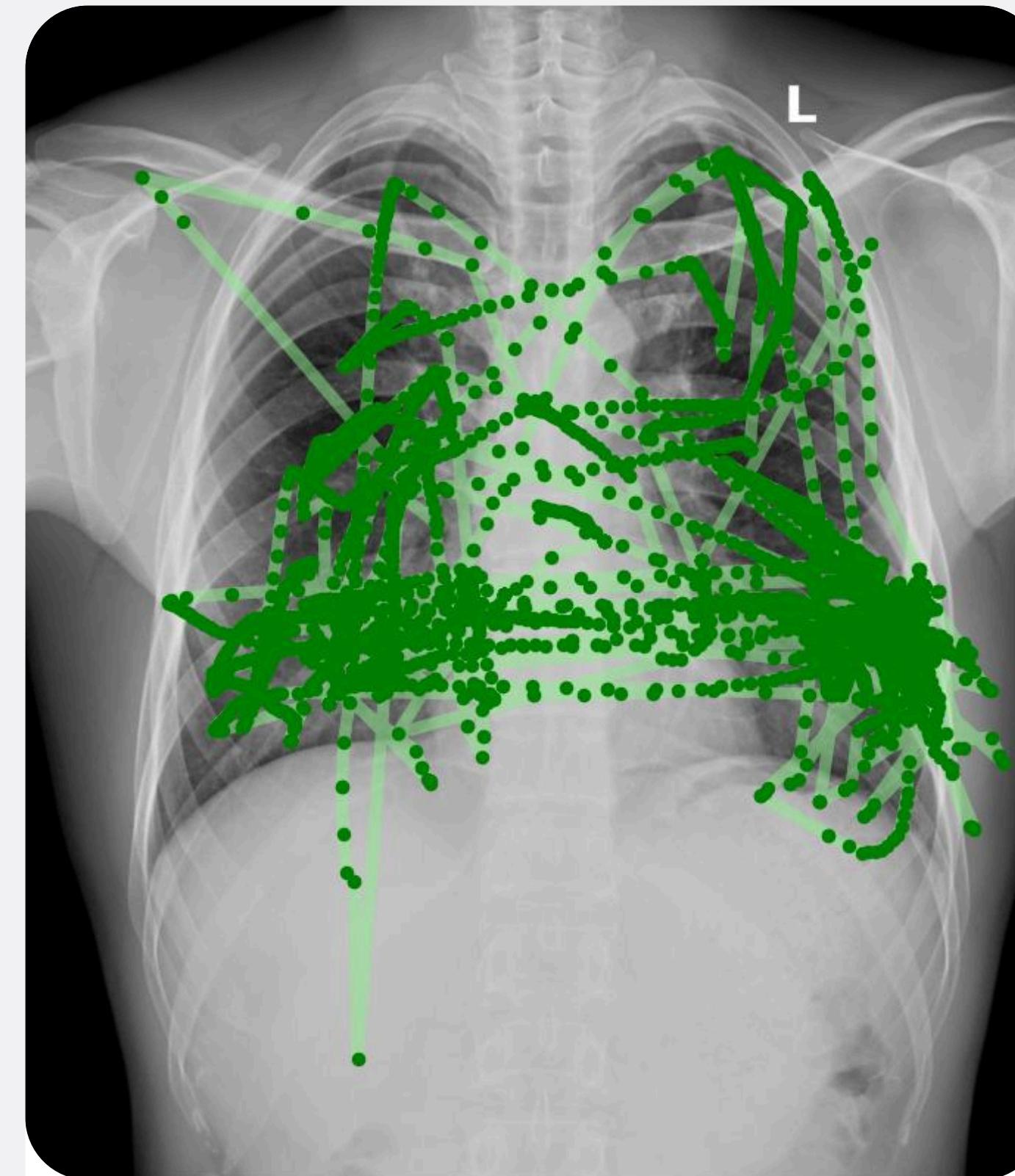
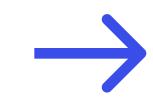
Математика в медицине  
Сеченовский университет, 2025

# ВЗГЛЯД ВРАЧА

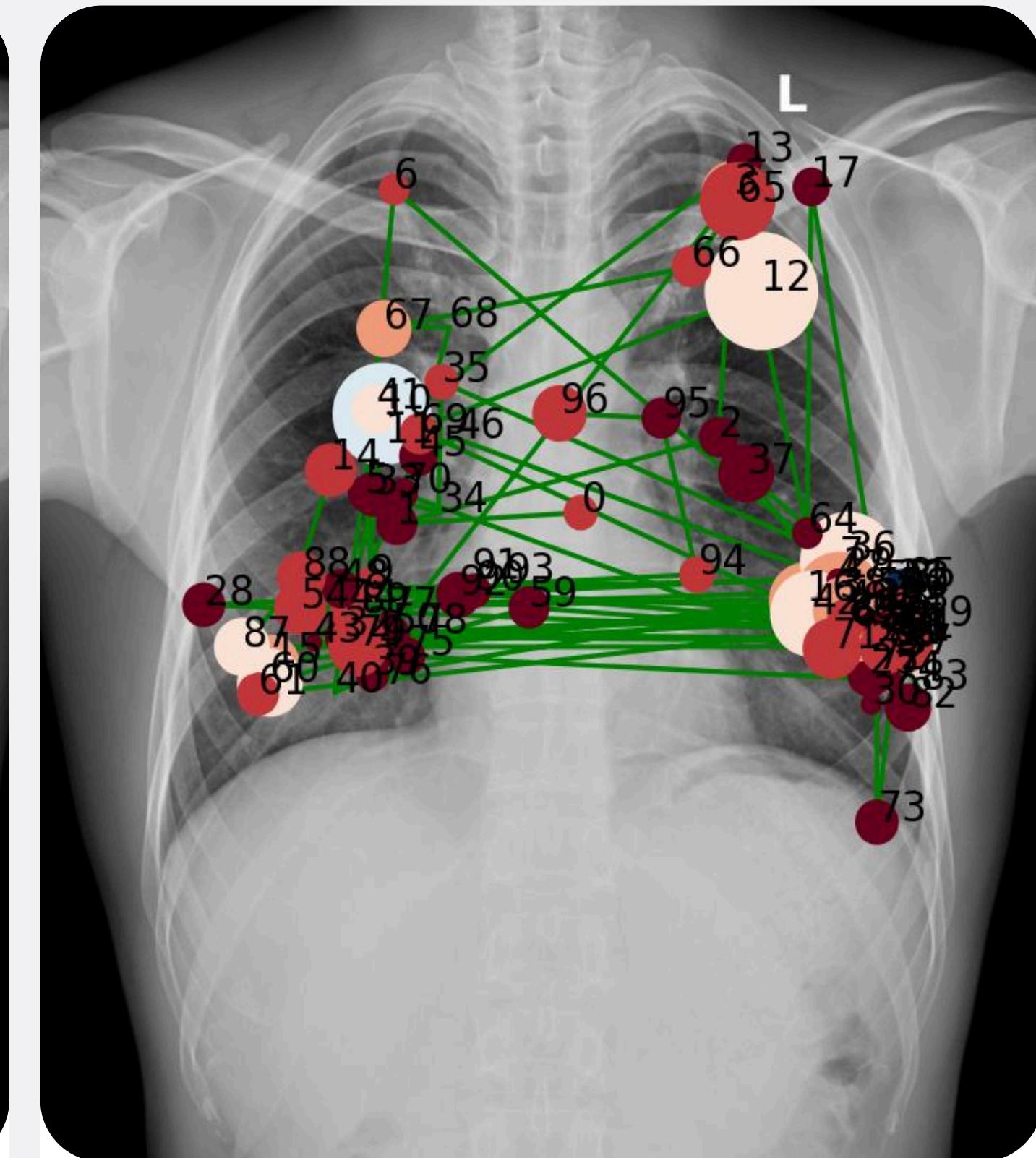
1



Анализ рентгена



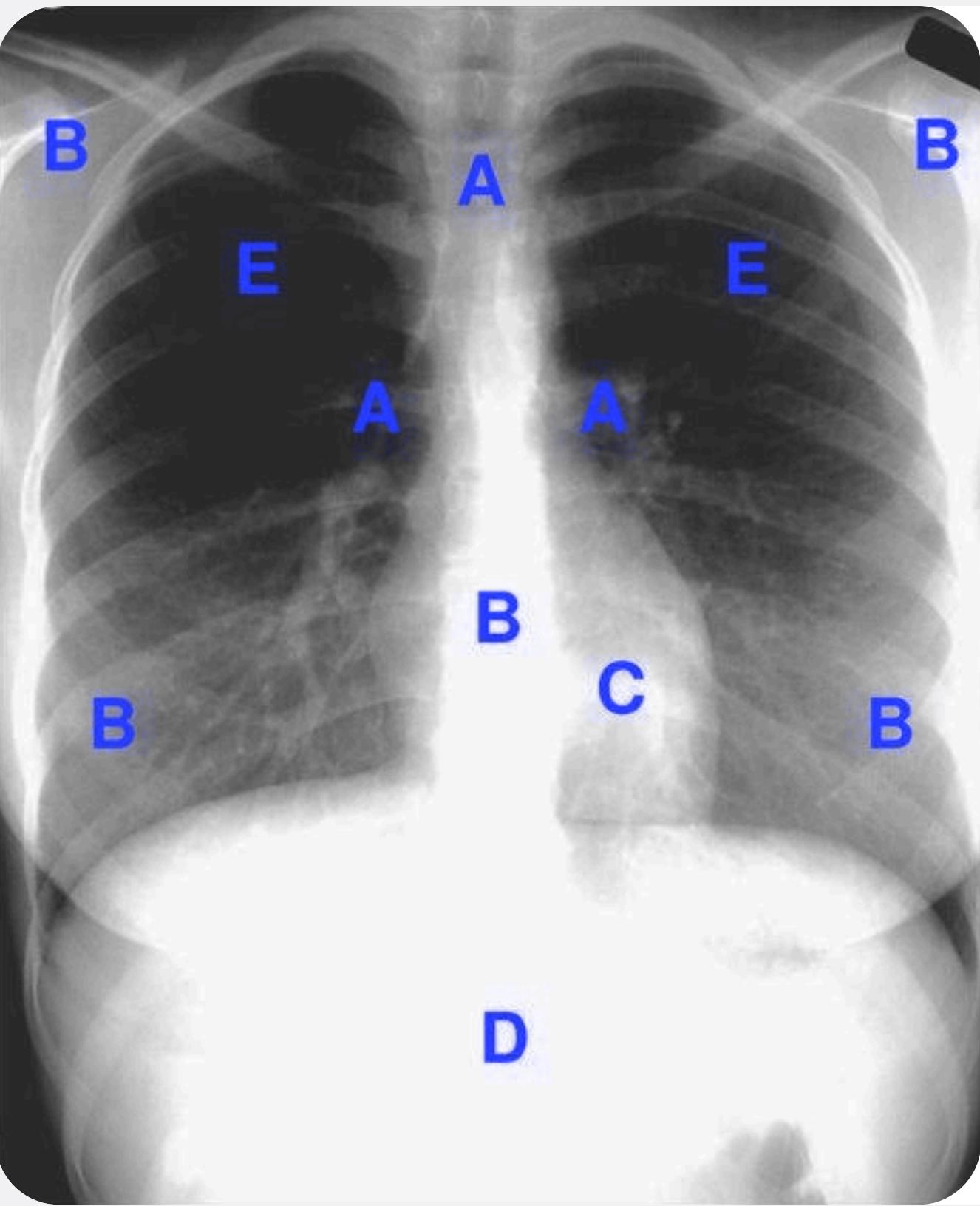
Исходные точки  
взгляда



Точки фиксации  
и саккад

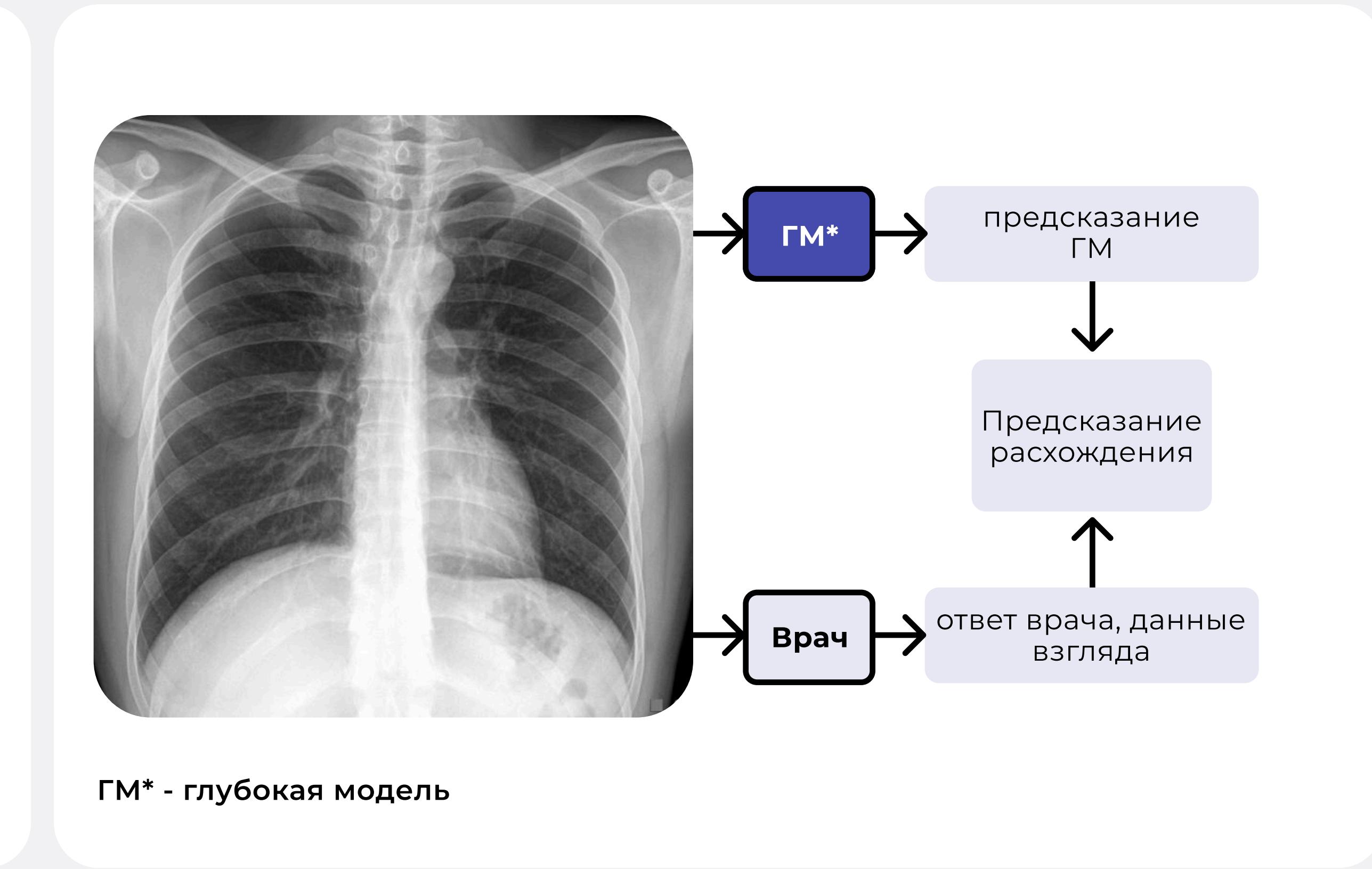
# ПРИМЕНЕНИЕ ВЗГЛЯДА

2



## Протокол ABCDE

- A** (airway) — воздушные пути
- B** (bones) — кости
- C** (cardiac) — сердце
- D** (diaphragm) — диафрагма
- E** (everything else = lungs)  
— все остальное = лёгкие



## Наборы данных со взглядом на рентгеновских снимках грудной клетки:

**Eye-Gaze** [2021] - 1083 снимка

**REFLACX** [2021] - 2616 снимков

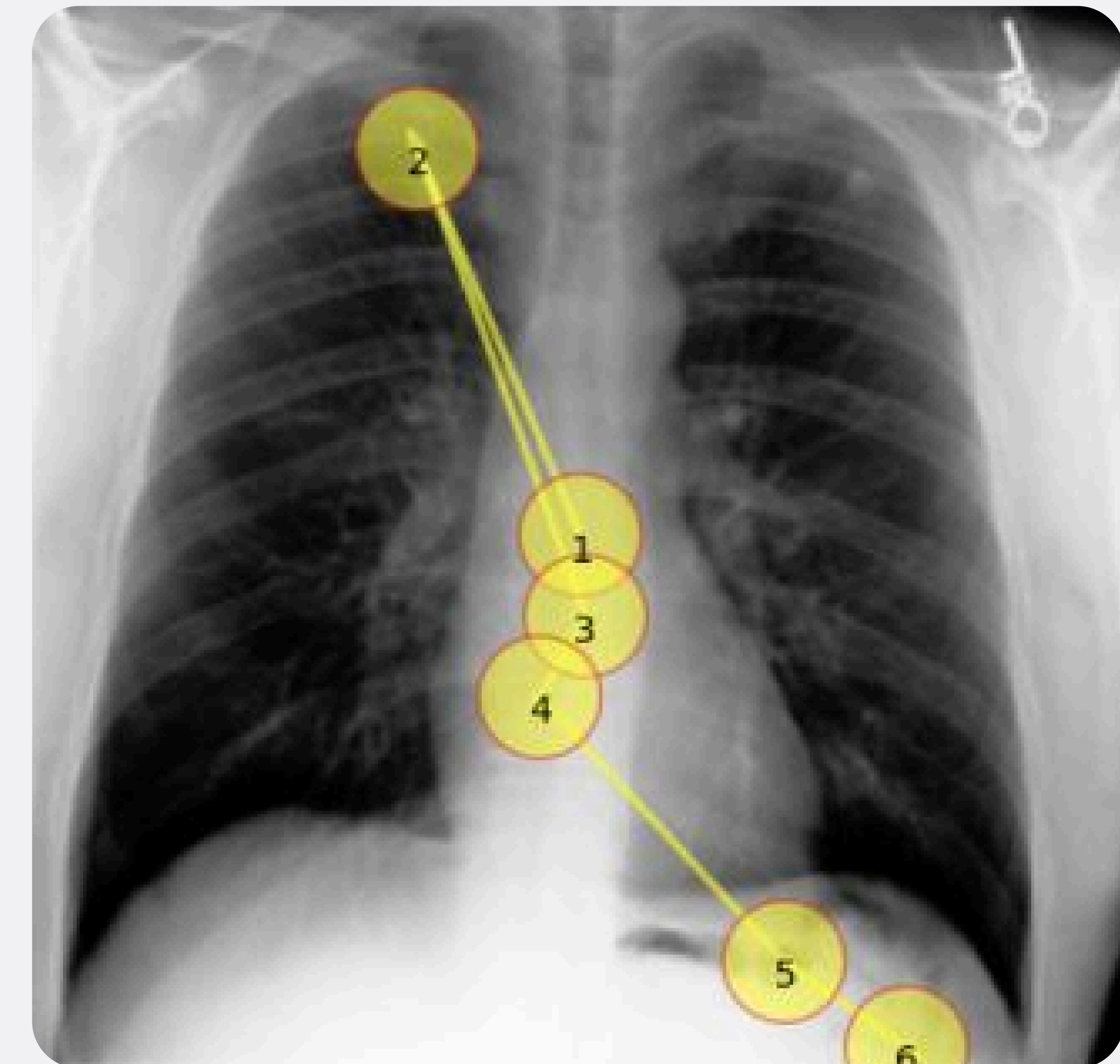
**FG-CXR** [2024] - 2951 снимков

**Итого:** 6 650 снимков

## ЗАДАЧА 1.

ПРЕДСКАЗАНИЕ ТРАЕКТОРИИ  
ВЗГЛЯДА ЧЕЛОВЕКА НА  
ИЗОБРАЖЕНИИ

3



# ЗАДАЧА 1. НАБОРЫ ДАННЫХ И МОДЕЛИ ЦЕЛЕНАПРАВЛЕННОГО ВЗГЛЯДА

4

## Наборы данных

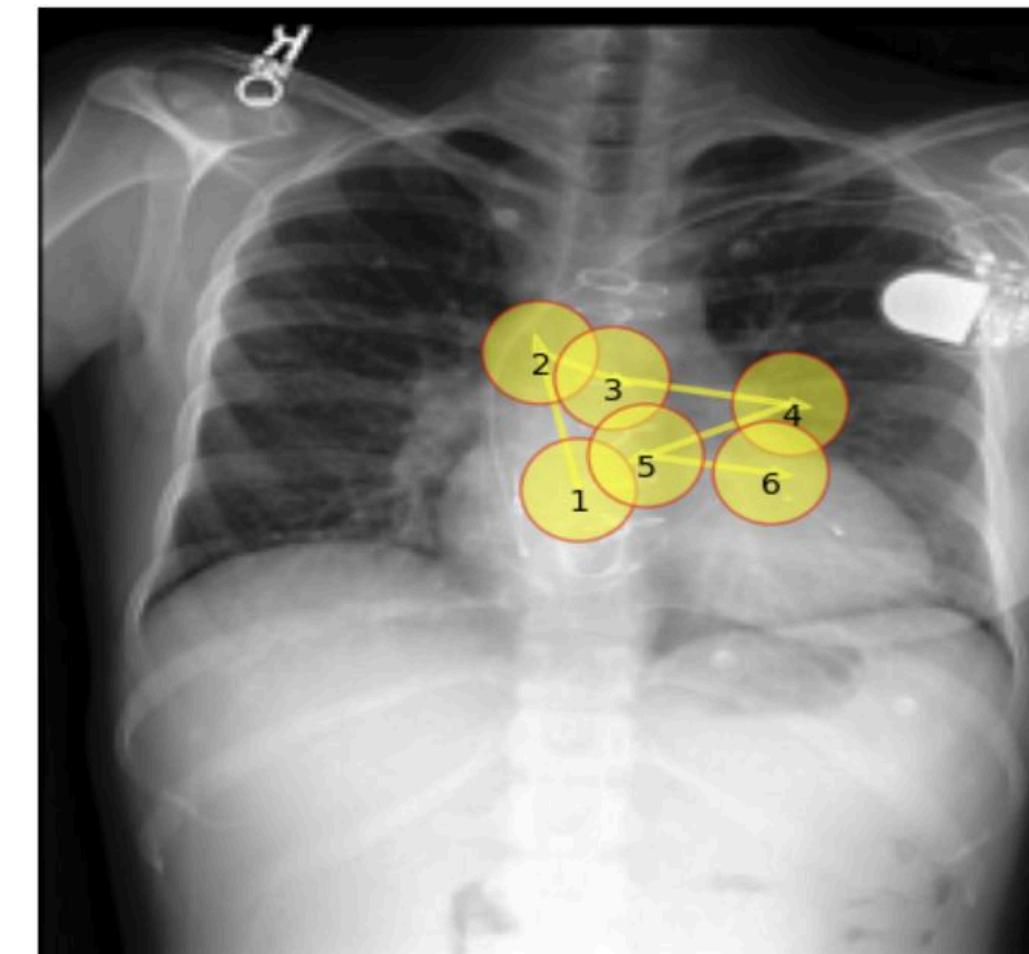
Объект из **COCO-Search18**.  
Цель — автомобиль.



**COCO-Search18** содержит 6202 изображений с разбиением:

- обучение - 70%, 4341
- валидация - 10%, 620
- тестирование - 20%, 1240

Объект из **GazeSearch**.  
Цель - кардиомегалия.



**GazeSearch** содержит 2081 изображения с разбиением:

- обучение - 70%, 1457
- валидация - 10%, 208
- тестирование - 20%, 416

## SOTA модели

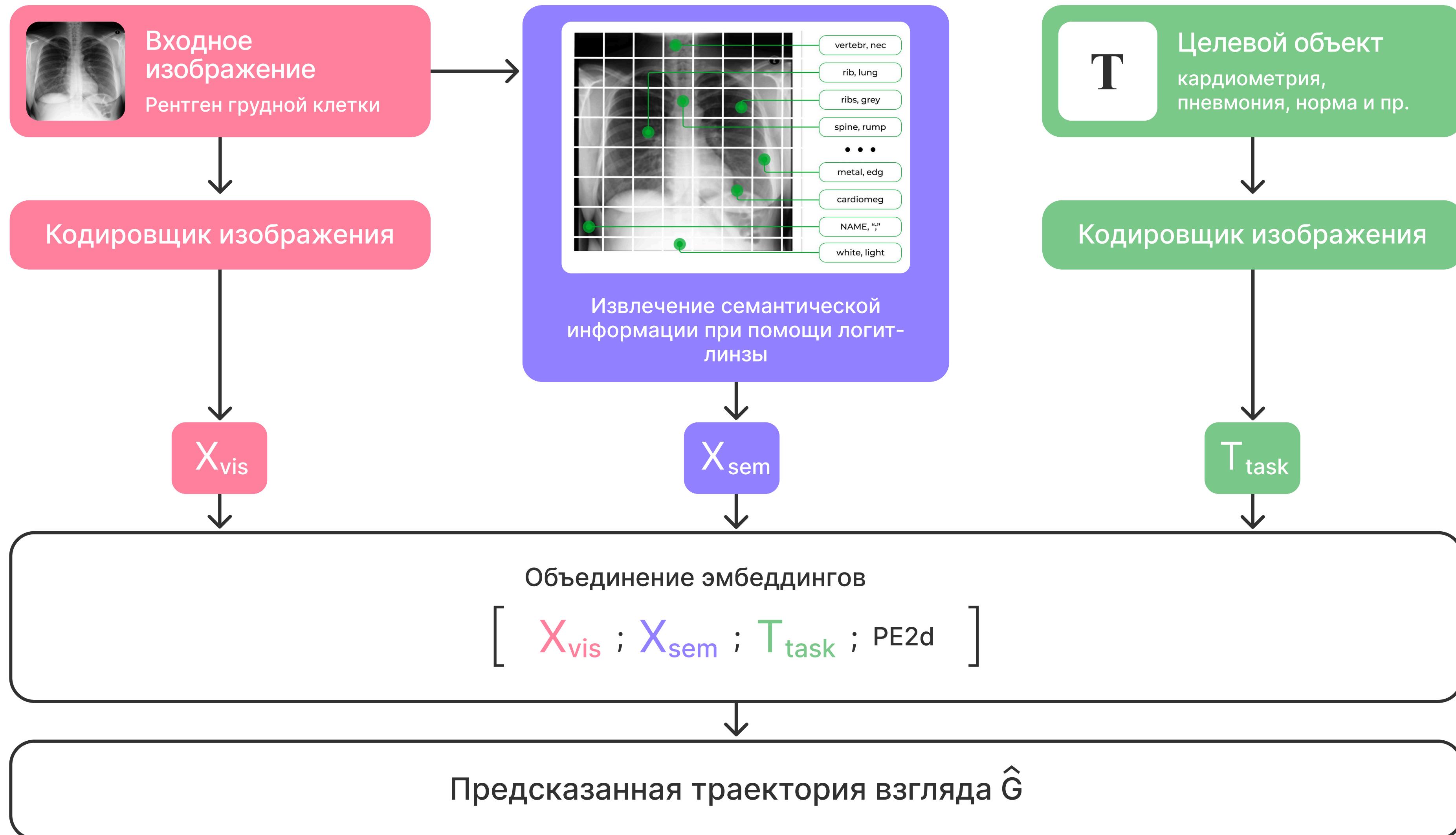
**GazeFormer** [2023] - COCO-Search18  
**HAT** [2023] - COCO-Search18  
**GazeSearch** [2024] - GazeSearch (рентгены)

В [1] адаптировали метод логит-линзы для VLM.

[1] C. Neo, L. Ong, P. Torr, M. Geva, D. Krueger, and F. Barez, 'Towards interpreting visual information processing in vision-language models', ICLR 2025.

# ЗАДАЧА 1. АРХИТЕКТУРА LOGITGAZE

5



# ЗАДАЧА 1. РЕЗУЛЬТАТЫ

6

## Предположения

- 1. Использование результатов логит-линзы в качестве доп. входа для **GazeFormer** улучшит качество на **COCO18-search**
- 2. Применение логит-линзы для **LLaVa-Med** позволит добиться лучших результатов на **GazeSearch**, чем с LLaVa
- 3. Использование медицинского кодировщика изображений позволит повысить качество результатов на **GazeSearch**

## Результаты

Значения метрик схожести траекторий взгляда (чем выше, тем лучше для ScanMatch/STDE, чем ниже, тем лучше для SED) на **GazeSearch**.

Метод	ScanMatch ↑		SED ↓	STDE ↑
	w/o Dur.	w/ Dur.		
GazeFormer	0.293 ± 0.021	0.201 ± 0.015	5.11 ± 0.08	0.799 ± 0.004
HAT	0.309 ± 0.020	–	5.07 ± 0.07	0.800 ± 0.004
GazeSearch	0.332 ± 0.019	0.223 ± 0.014	4.88 ± 0.06	0.809 ± 0.004
LogitGaze	0.328 ± 0.018	0.225 ± 0.015	5.07 ± 0.07	0.810 ± 0.004
LogitGaze-Med (Res)	0.416 ± 0.017	0.325 ± 0.012	4.68 ± 0.05	0.852 ± 0.003
<b>LogitGaze-Med (CheX)</b>	<b>0.419 ± 0.016</b>	<b>0.330 ± 0.010</b>	<b>4.68 ± 0.05</b>	<b>0.855 ± 0.003</b>

Тестирование **LogitGaze** на наборе данных **COCO-Search18**.

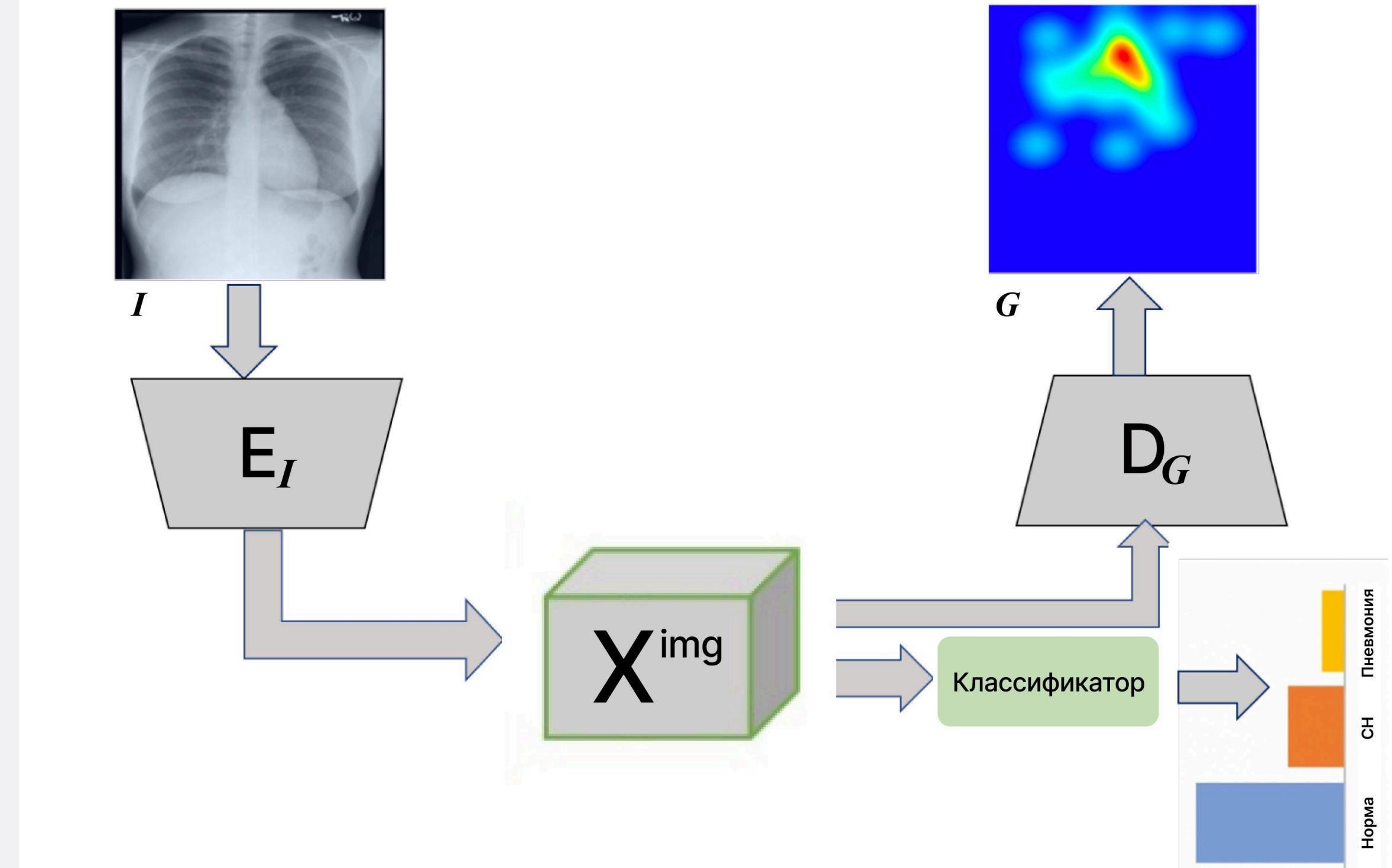
Метод	ScanMatch	
	w/o Dur	w/ Dur
LogitGaze	<b>0.527 ± 0.012</b>	<b>0.454 ± 0.013</b>
GazeFormer [48]	0.492 ± 0.014	0.441 ± 0.015
LogitGaze-Med	0.353 ± 0.018	0.289 ± 0.020

**Вывод:** SOTA результат на медицинском и общем домене.

## ЗАДАЧА 2.

7

### ВНЕДРЕНИЕ ВЗГЛЯДА РЕНТГЕНОЛОГА В ГЛУБОКУЮ МОДЕЛЬ КЛАССИФИКАЦИИ ПАТОЛОГИЙ ЛЕГКИХ



# ЗАДАЧА 2. ДАННЫЕ, МЕТОД И РЕЗУЛЬТАТЫ

8

## Предположения

В [1] представили

- набор данных взгляда (1083 снимка);
- метод для улучшения классификации при помощи взгляда рентгенолога

1. Замена натуральных данных взгляда на синтетические в методе [1] позволит еще сильнее улучшить результаты

2. Использование медицинской модели **LogitGaze-Med** для генерации взгляда лучше, чем **LogitGaze**

## Метод проверки

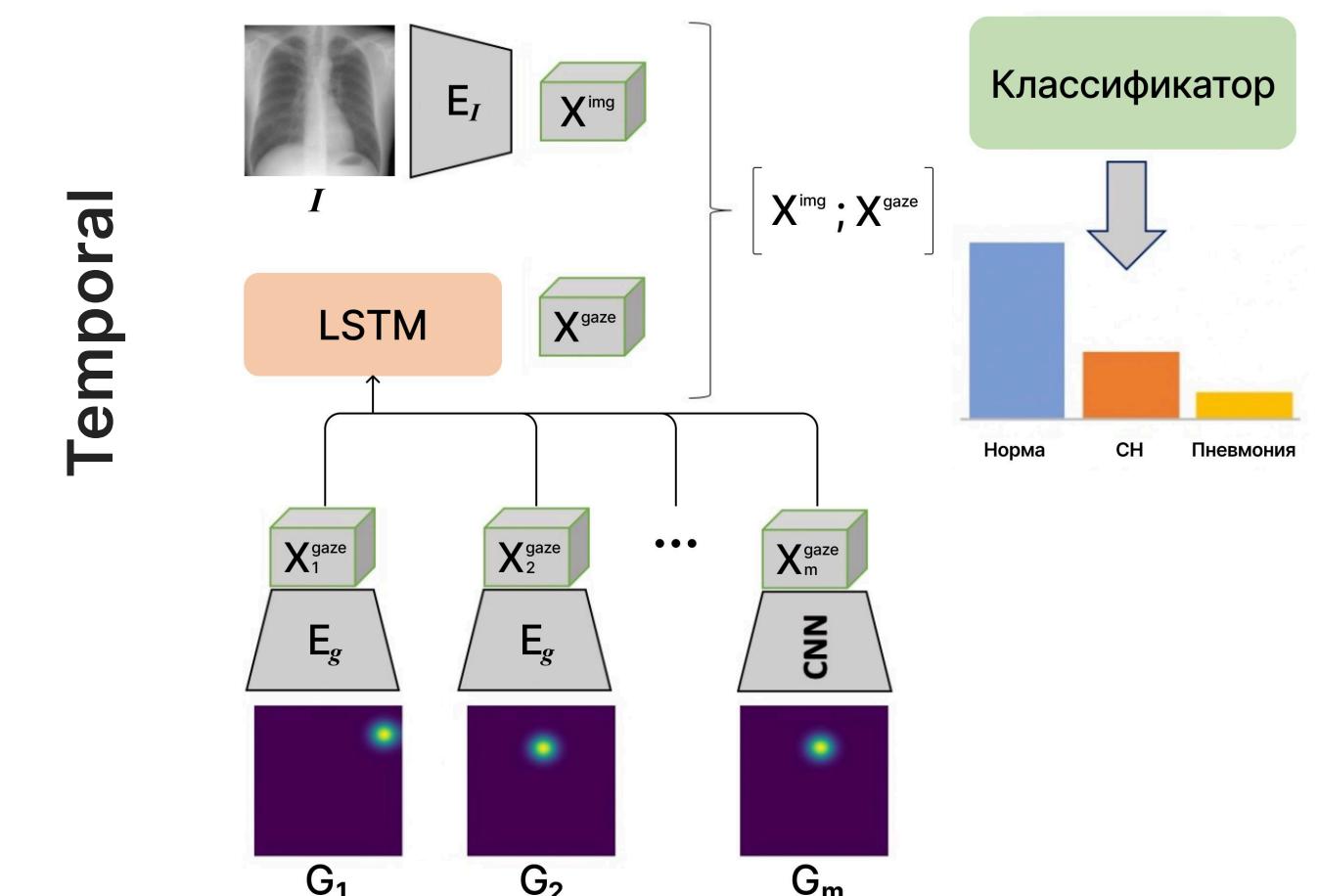
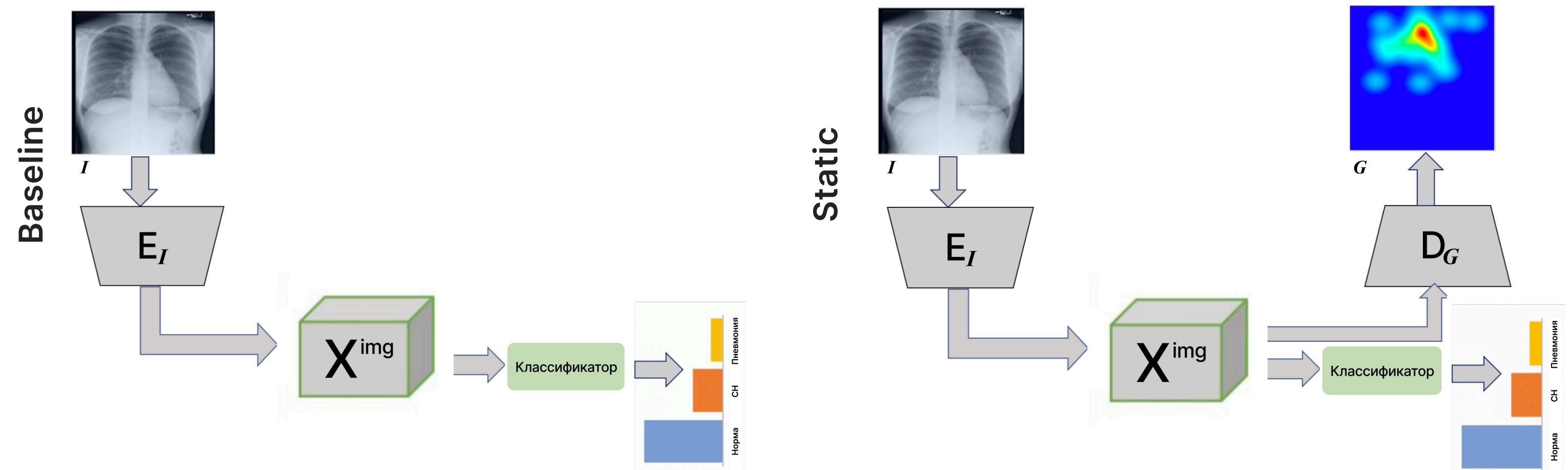
- Создать синтетический набор данных взгляда (**LogitGaze-Med**, **LogitGaze**, **GazeFormer**) на основе **MIMIC-CXR** (227 тыс. рентгенов);
- Использовать синтетический набор как тренировочный в эксперименте [1];
- Провести тестирование на тестовом наборе данных из [1].

## Результаты

Среднее значение AUROC по патологиям: норма, застойная сердечная недостаточность и пневмония).

Взгляд	Baseline	Temporal	Static
Eye-Gaze [11]	$0.77 \pm 0.02$	$0.82 \pm 0.03$	$0.87 \pm 0.02$
GazeFormer [9]	$0.78 \pm 0.02$	$0.84 \pm 0.02$	$0.89 \pm 0.01$
LogitGaze	$0.80 \pm 0.01$	$0.87 \pm 0.02$	$0.90 \pm 0.01$
<b>LogitGaze-Med</b>	$0.82 \pm 0.01$	$0.90 \pm 0.02$	$0.91 \pm 0.01$

[1] A. Karargyris et al., "Creation and validation of a chest X-ray dataset with eye-tracking and report dictation for AI development," Sci Data, vol. 8, no. 1, Mar. 2021, doi: [10.1038/s41597-021-00863-5](https://doi.org/10.1038/s41597-021-00863-5).



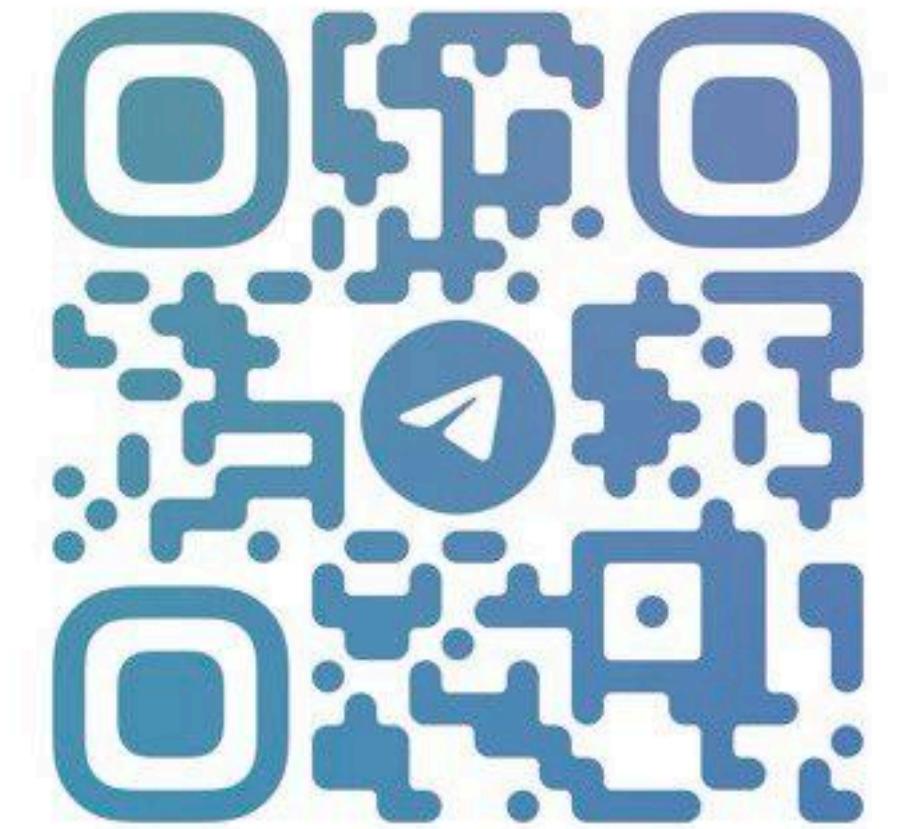
Результаты приняты на main track NeurIPS 2025 (A\*).

# СПАСИБО ЗА ВНИМАНИЕ!

Илья Першин

Университет Иннополис

[i.pershin@innopolis.ru](mailto:i.pershin@innopolis.ru)



@STRAID