



**ПОЛИТЕХ**

Санкт-Петербургский  
политехнический университет  
Петра Великого

COMPUTO ERGO SUM

***XXXIV МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ «МАТЕМАТИЧЕСКИЕ МЕТОДЫ В  
ТЕХНИКЕ И ТЕХНОЛОГИЯХ – ММТТ-34  
ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ СУПЕРКОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ С  
ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ***

26 октября 2021 г.  
г. Минск

Заборовский В.С. , Л. В. Уткин  
Политехнический университет  
Петра Великого, Санкт-Петербург

# ОСНОВНОЙ ВОПРОС «ЦИФРОВОЙ ТРАНСФОРМАЦИИ»

## ИСТИННО ЛИШЬ ТО, ЧТО МОЖНО

- Вычислить на компьютере (? теоремы Геделя) а полученный результат
- Понять и объяснить ( ? неформализуемость истны)

*В чем суть алгоритмической вычислимости и как ее достичь, используя феномен ИИ (....) ?*

1) **Вычислимость** – описание решения **«ПРЯМЫХ»** задач с помощью программ-алгоритмов, реализующих счетный набор частично рекурсивных операций ?

2) **Интеллектуальность** - решение **«ОБРАТНЫХ»** задач путем построения АЛГОРИТМОВ вычислений и объяснения (интерпретации) результатов

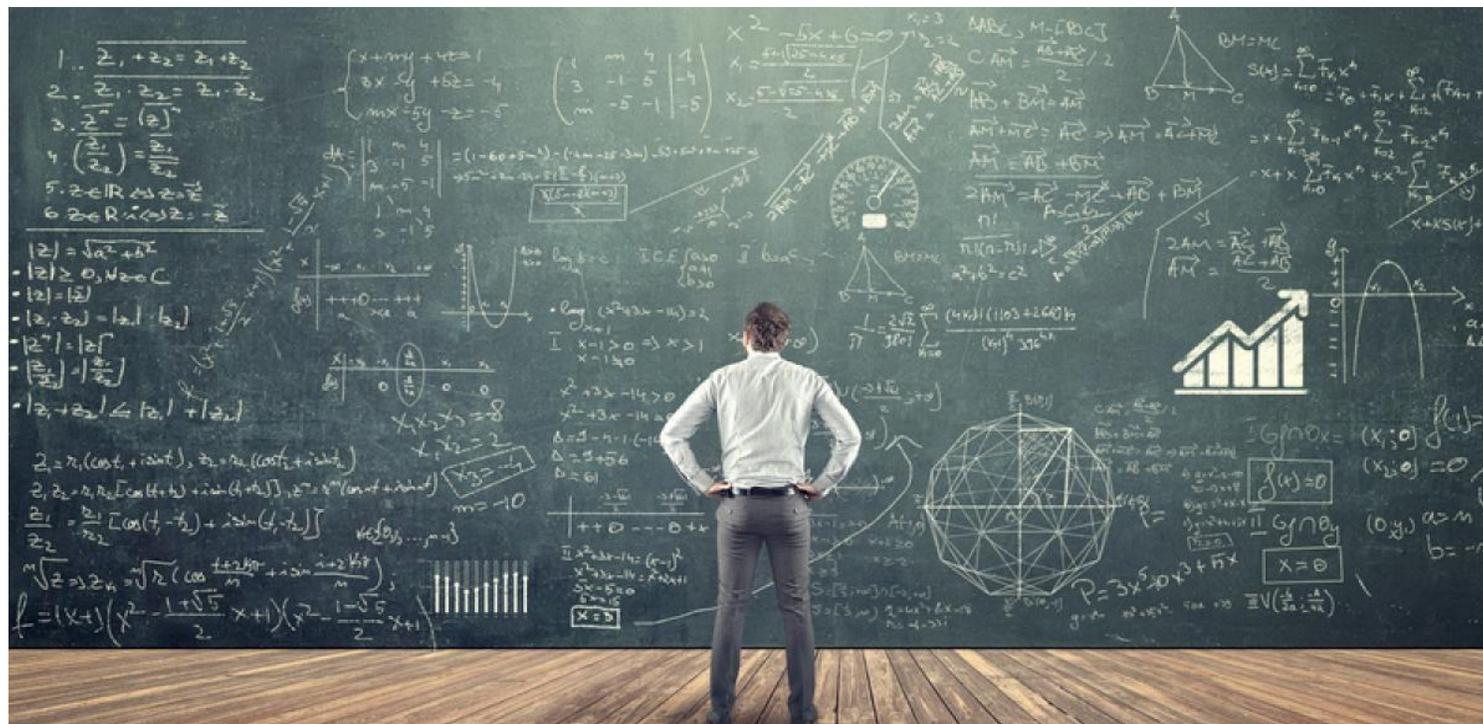
**Интерпретируемость** (объяснение) результатов - это степень уверенности ... в том, что вычисленный результат можно **ПОНЯТЬ**, а используемый алгоритм вычислений **УТОЧНИТЬ** так, чтобы повысить степень уверенности и снизить «коэффициент незнания»

## Для ответа на основной вопрос...

**Важно понимать:**  
каждая **вычислимая**  
функция  $f(x)$  может быть  
представлена **счетным**  
множеством различных  
программ!

**и уметь отвечать**  
**на вопросы:**  
Какую программу из  
возможных следует  
**выбрать** (написать),  
чтобы **понять**  
**(объяснить)** полученные  
результаты ?

Если вы не можете **объяснить** что-либо  
простыми словами, вы это не **понимаете**  
Р. Фейнман

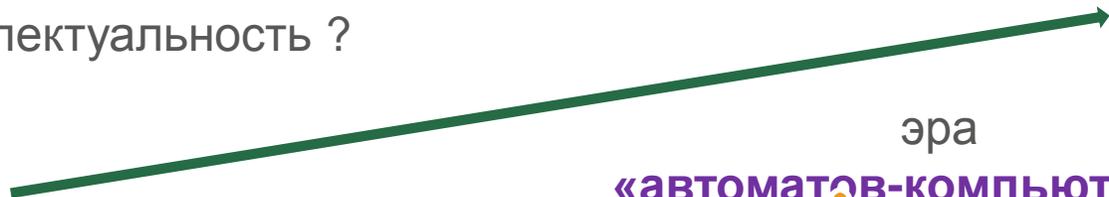




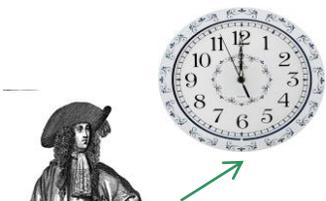
Физика достигла таких высот, что мы можем рассчитать даже то, что невозможно себе представить

Л. Ландау

За счет чего повышается интеллектуальность ?



Эра часов автоматов, управляемых одной программной



«сила» мысли - алгоритм



механическая энергия-вычисления



эра «арифмометров», управляемых человеком 1900-1960



программа-алгоритм

результат-объяснение



1960 – 2020

эра «автоматов-компьютеров», вычисляющих решения под управлением программ

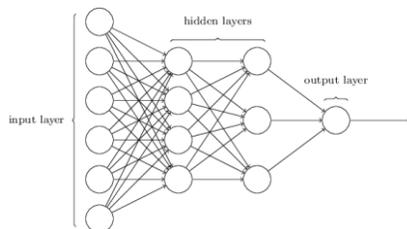


Эл. энергия-вычисления

эра компьютерных систем, вычисляющих решения под управлением данных и «объясняющих» программ 2020 >

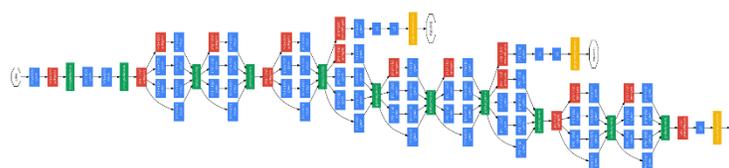


данные-программа вычислений



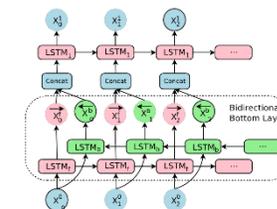
## Многоуровневый перцептрон

- классификация
- функциональная аппроксимация
- автокодер



## Сверточные нейронные сети

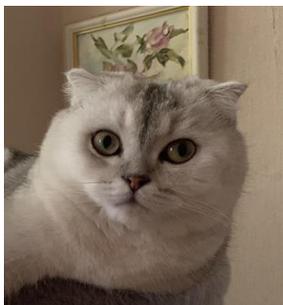
- выделение свойств
- детектирование объектов
- сегментация изображений



## Рекуррентные нейронные сети

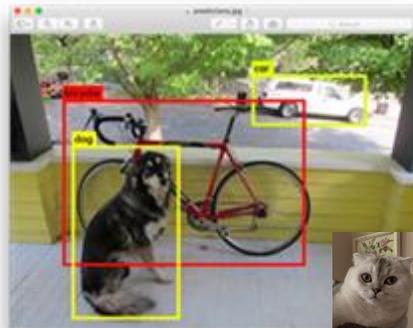
- Обработка временных потоков
- Преобразования разговора в текст
- Перевод текстов

## классификация



“КОТ”

## Обнаружение объектов



## Сегментация данных

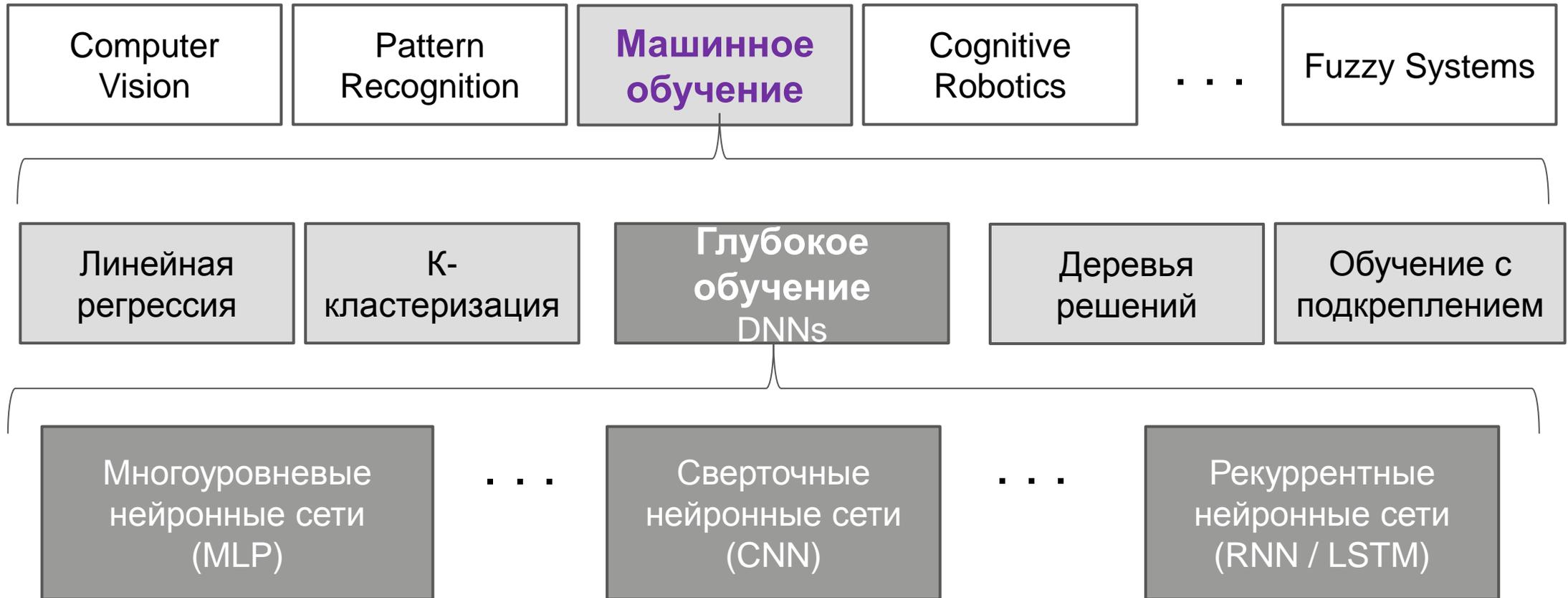




ПОЛИТЕХ

# ПЛАТФОРМЫ ИНТЕЛЛЕКТУАЛИЗАЦИИ ВЫЧИСЛЕНИЙ

Используя платформы PyTorch и Tensorflow, можно распределить рабочую нагрузку по сотням (или тысячам) идентичных GPU (должны иметь одинаковую память и вычислительную производительность) **время обучения ИНС можно сократить**

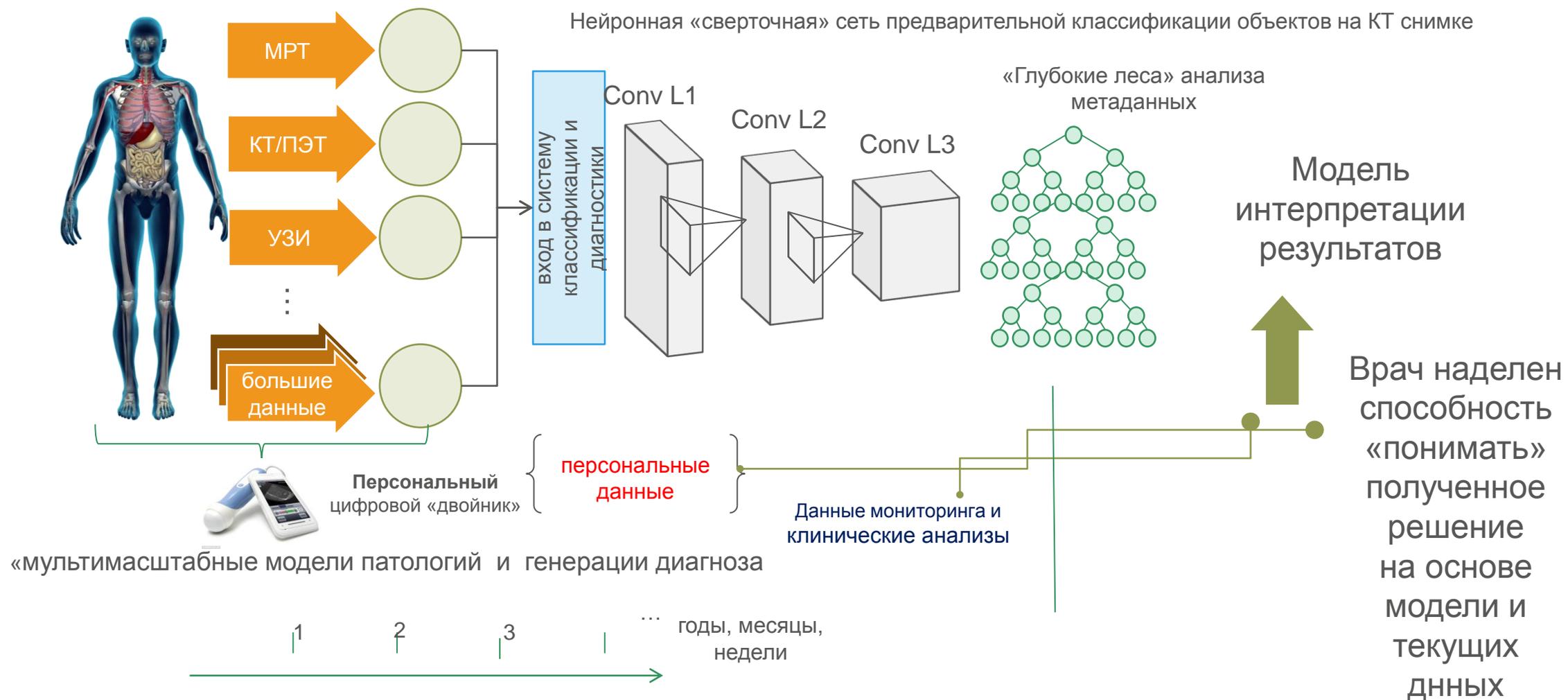


Платформы существуют, но: трудоемкость обучения высокая, а робастность результатов низкая

# ПРИМЕР «ЦИФРОВОЙ ДВОЙНИК ПАЦИЕНТА»: АЛГОРИТМИЧЕСКАЯ МОДЕЛЬ + ТЕКУЩИЕ ДАННЫЕ

Потоки первичных данных, характеризующих состояние биологических тканей и процессов

Суперкомпьютерный центр хранения, анализа и предсказательной диагностики



Пример: задача обучения ГЛУБОКОЙ ИНС распознавать 3D изображения компьютерной томографии (КТ)



СК: 1Пфлопс  
=  $1 \times 10^{15}$  Флопс

Для обучения н/с с точностью классификации **90%** надо:

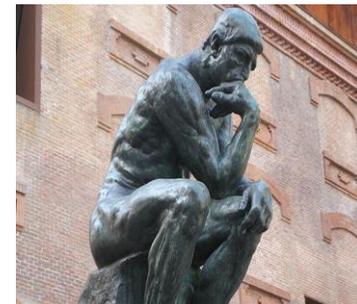
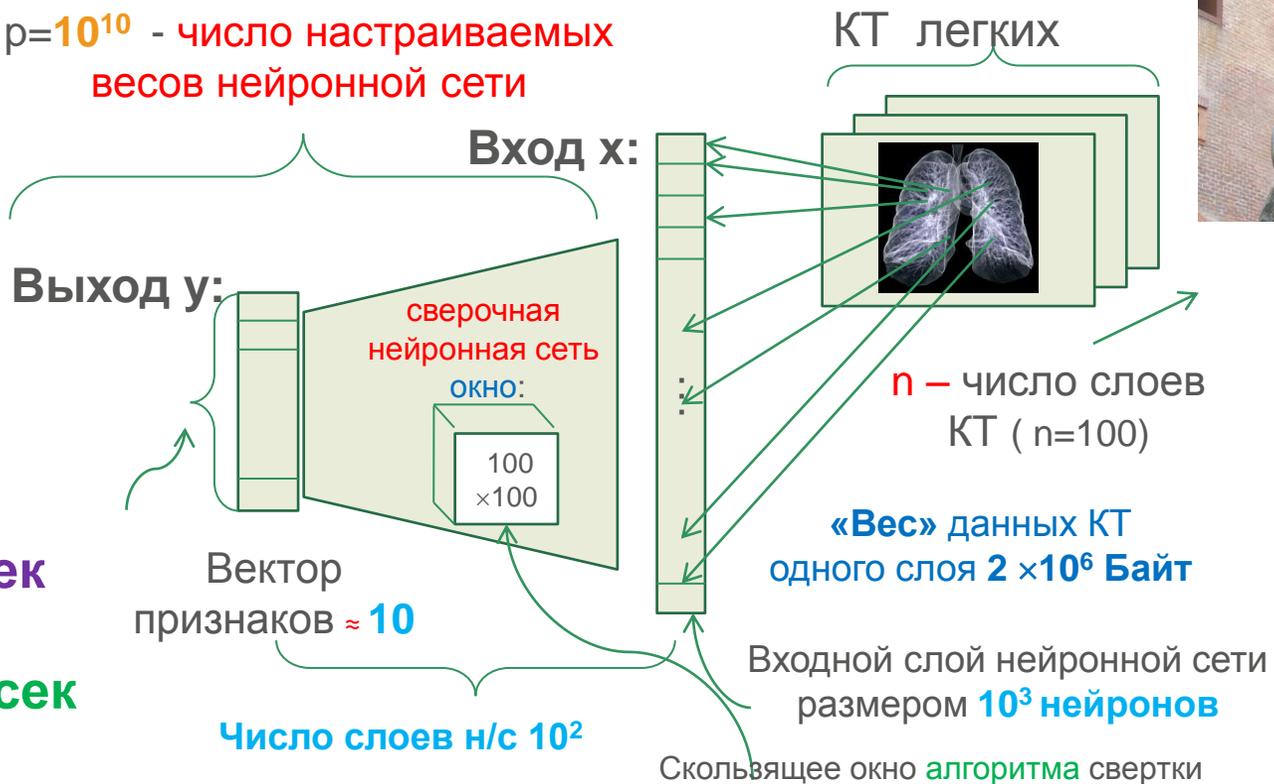
- Число операций  $\approx 10^{17}$
- Время обучения ИНС на СК  $\approx 10^2$  сек
- t обучения на ПК (200 ГФлопс)  $\approx 2 \times 10^6$  сек

**Градиентный алгоритм** обучения **глубокой сверточной нейронной сети**:

- функция ошибки  $F = \|y^* - y\|^2$ ,  $y^*$  - эталонный вектор признаков
- Алгоритм использует  $10^{10}$  частных производных  $F$  по всем настраиваемым параметрам;
- Число операций численного дифференцирования на одну итерацию  $Q = 10^{15}$

Обучающая выборка  **$10^4$  КТ снимков**

$p = 10^{10}$  - число настраиваемых весов нейронной сети

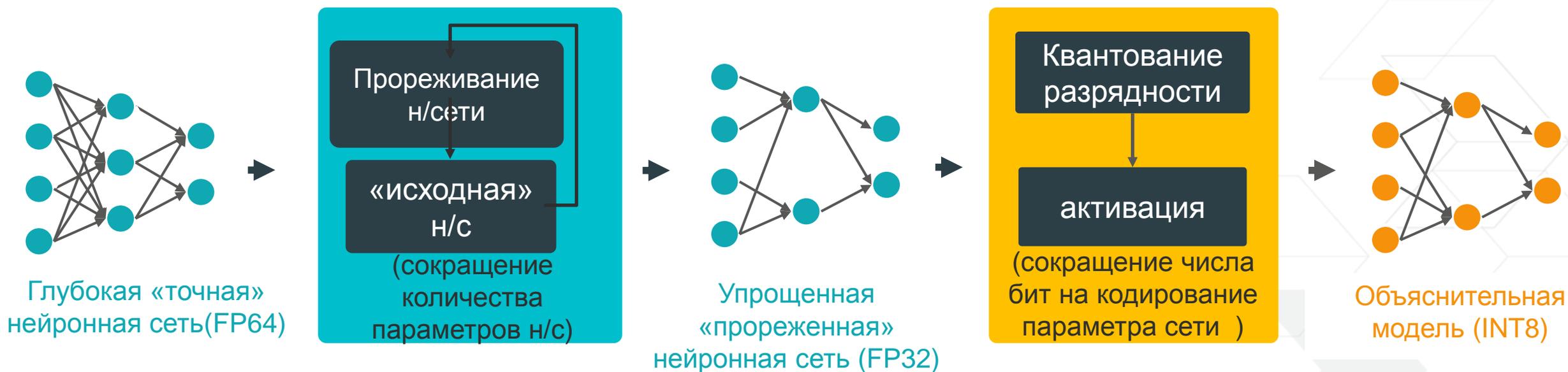


обучение ИНС даже среднего размера может занимать **недели** или **месяцы**



ПОЛИТЕХ

# ЧТО ВАЖНЕЕ – ТОЧНОСТЬ ИЛИ ИНТЕРПРЕТИРУЕМОСТЬ МОДЕЛИ: FP 64 vs INT 8



Процесс включает два инструментальных этапа

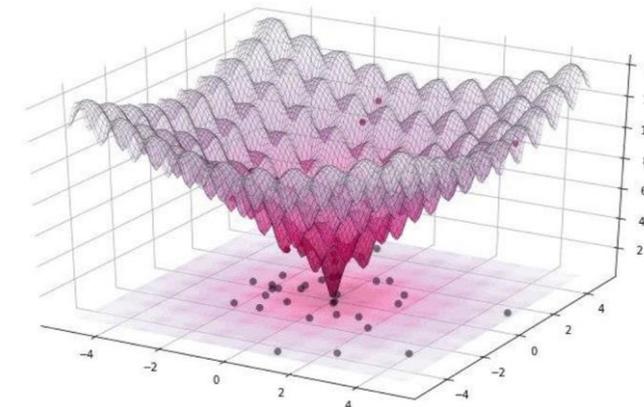
- Сокращение числа параметров н/с –  $p$  раз ?
- Сокращение разрядности данных –  $q$  раз ?

Эффекты

- Сжатие размера модели  $5x \sim 100x$
- Сокращение времени вычислений  $1.5x - 10x$

**Объяснительную модель можно 1) быстро посчитать; 2) результаты легко интерпретируются и 3) используется вместе с другими метриками, которые типичны для СКТ**

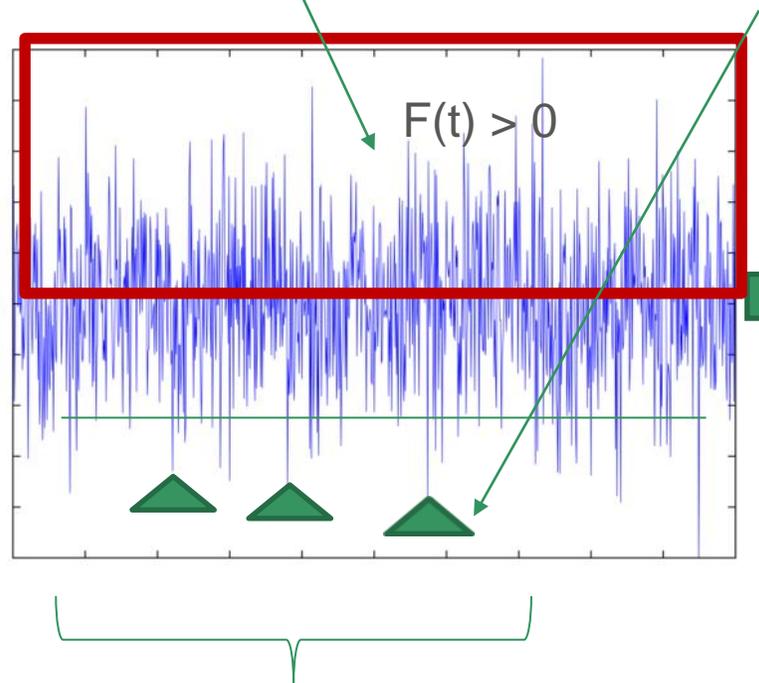
- Перечислимость,
  - вычислимость,
    - разрешимость ...
      - **Объяснимость (интерпретируемость)**



В силу **сложности** применяемых алгоритмов и гетерогенности вычислительных структур имеется потребность использовать мультимодальные метрики, которые позволяют определять какие:

- параметры модели следует учесть, чтобы **объяснить** результат ?
- характеристики программы вычислений **влиять** на точность результата ?

Сложный физический процесс можно моделировать в различные моменты времени



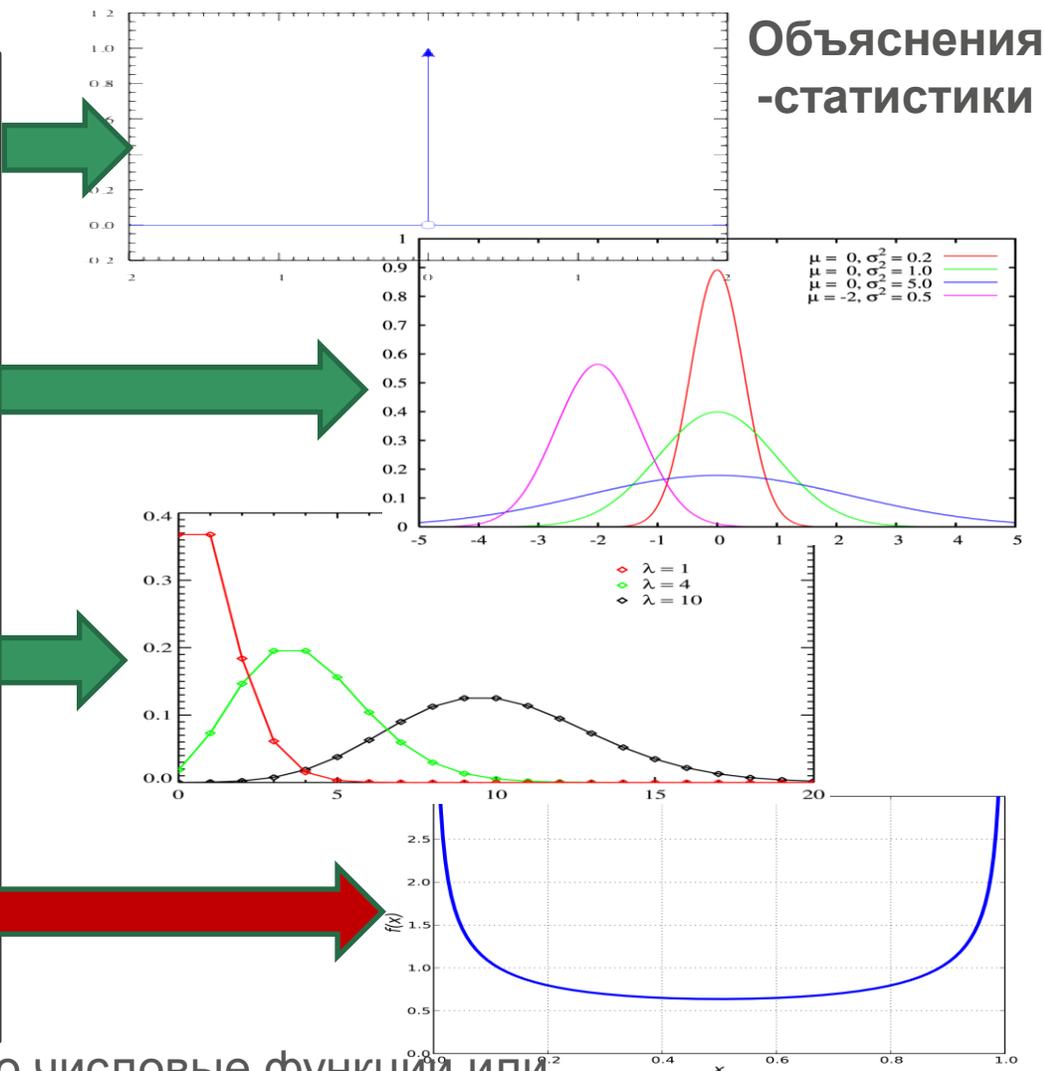
На конечных интервалах наблюдения сложные процессы могут иметь различные «модели объяснения»

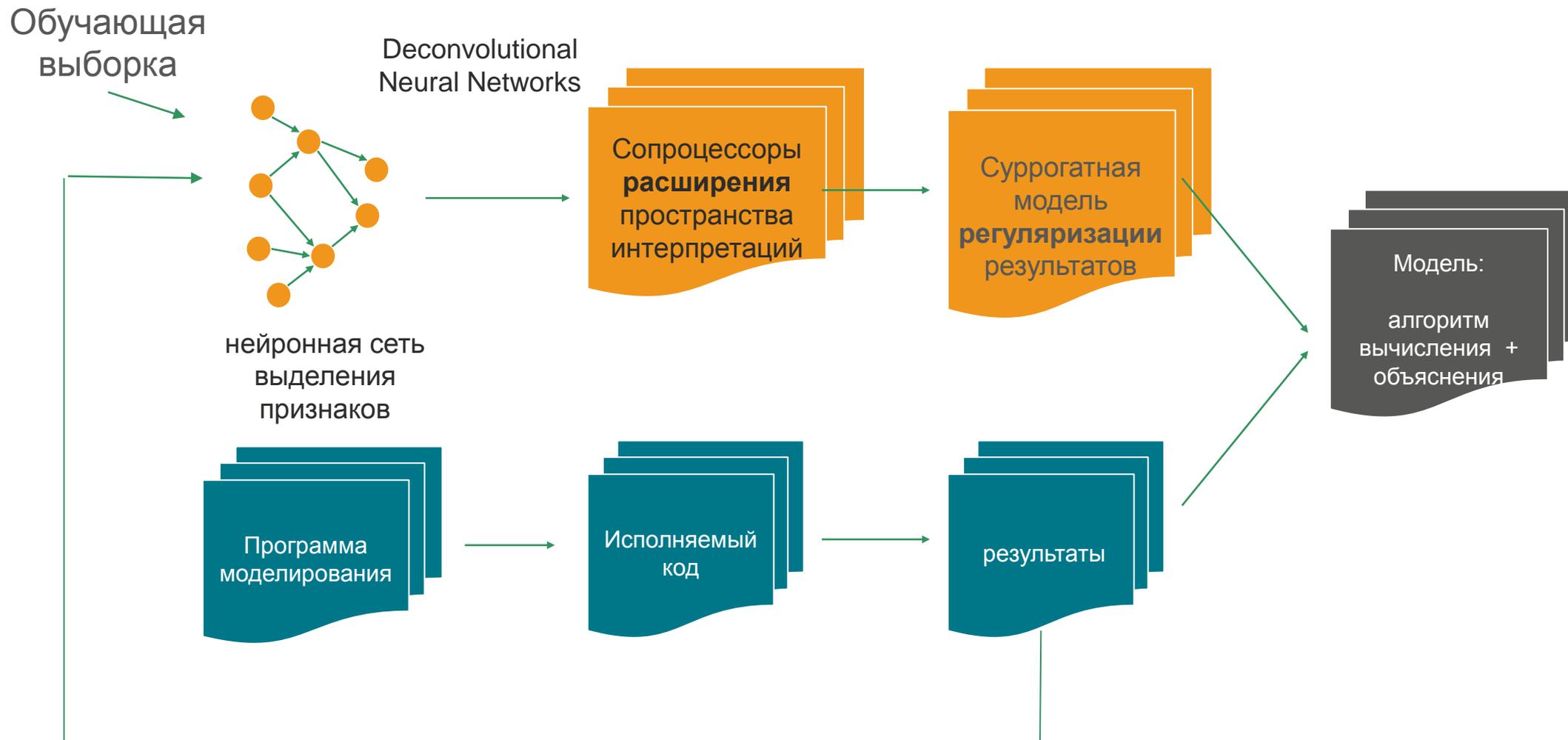
Интеллектуальный классификатор – видим то, что понимаем



Принятие решения, какими данными можно пренебречь

Вопрос: вычисления – это числовые функции или манипулирование символами (образами) ? .





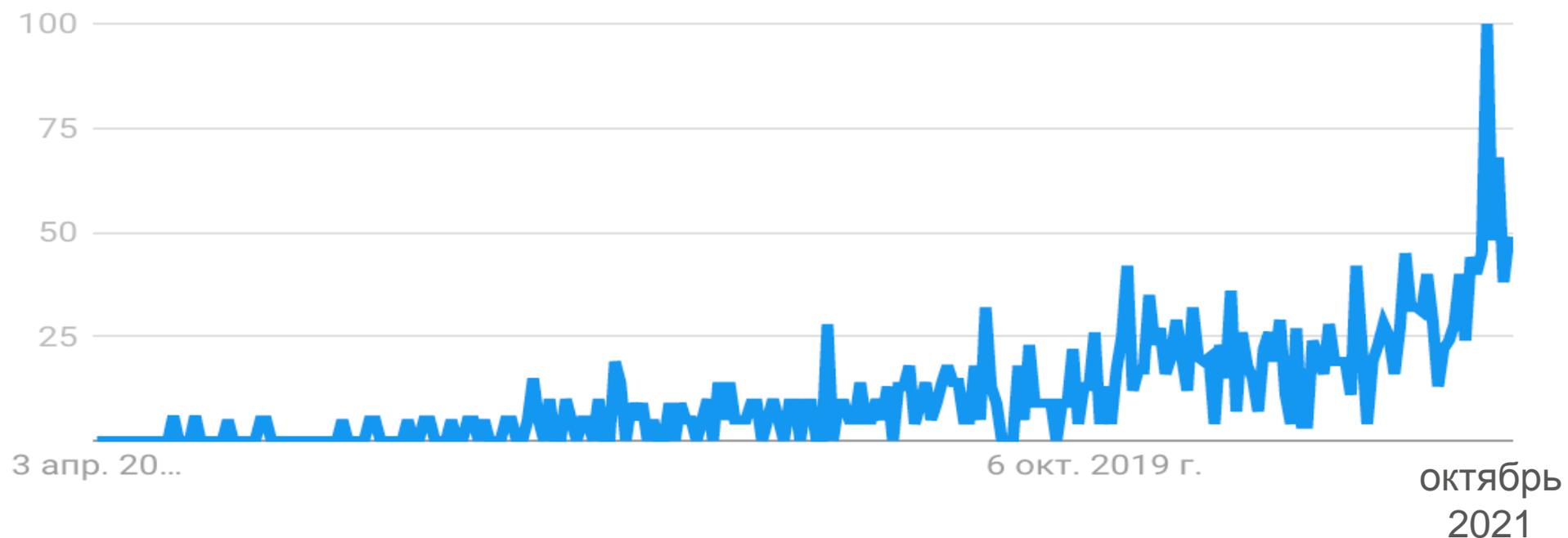
*Ключевые слова: объяснение, интерпретация, понимание, доверие*

- Дедуктивные **объяснения** (explanation) в терминах понятий
- Индуктивная **интерпретация** (interpretation) в терминах параметров
- **Понимание** сделанных предсказаний (understanding) например с помощью SHAP (SHapley Additive exPlanation)
- **Доверие** к результатам функционирования модели (trust)

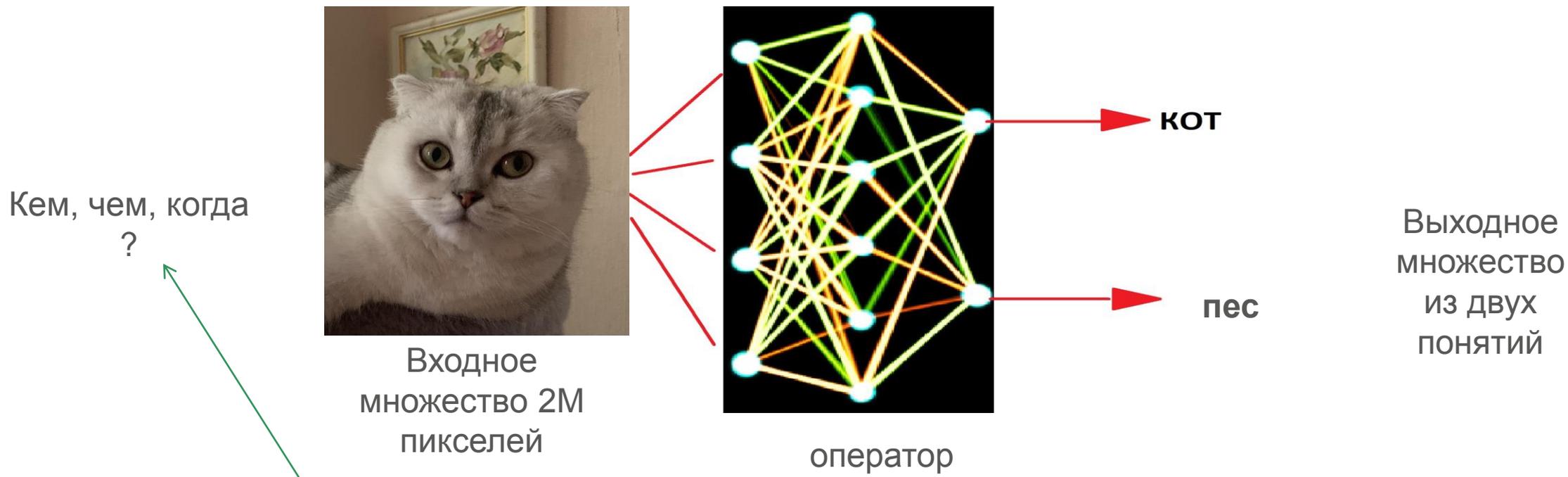
Исходим из того, что если пользователи не доверяют модели или прогнозу, то они не будут их использовать.

# Публикации на тему XAI

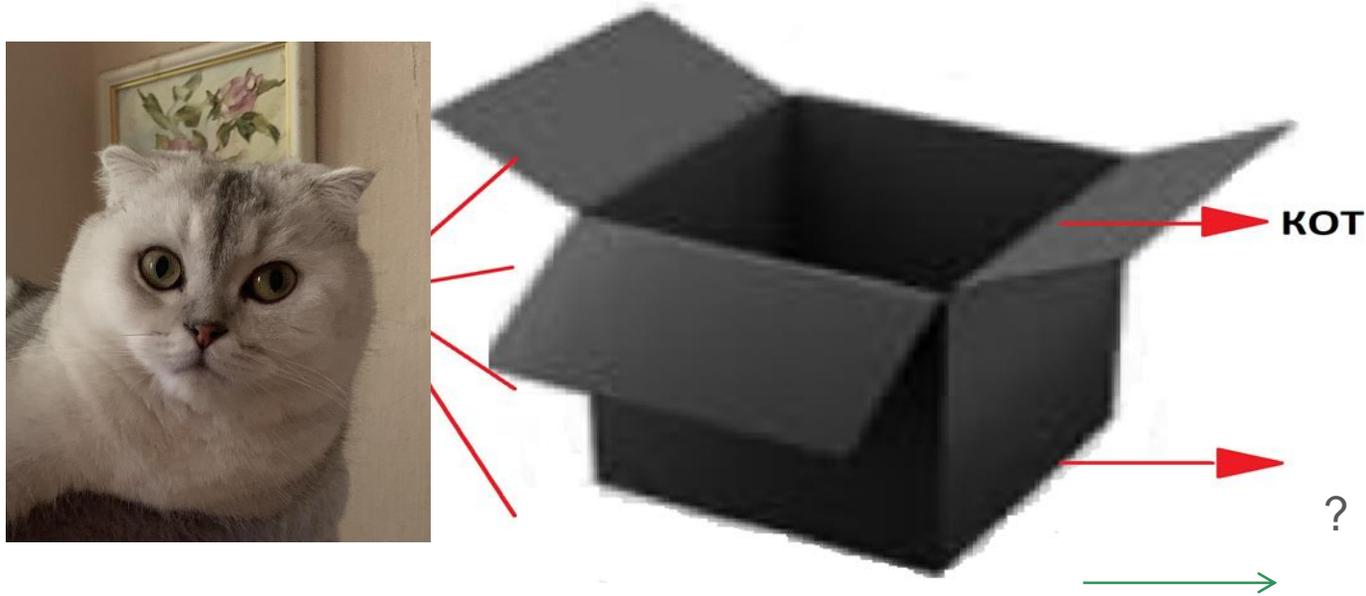
● Explainable AI



По всему миру. Последние 5 лет. Веб-поиск.



"заранее **обученная** нейроморфная модель предсказывает, что **2 М пикселей** входного изображения **кодируют** с помощью hidden **сигнатуры** в выходном embedding vector позицию «кот» с вероятностью **0.98**", а слово «пес» с вероятностью **2%**.

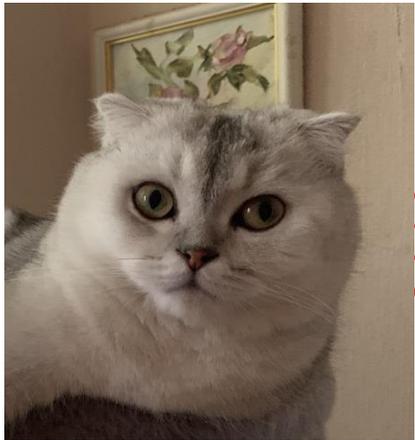


«черный ящик» предсказывает, что на фотографии изображен **КОТ** с вероятностью **0.98**

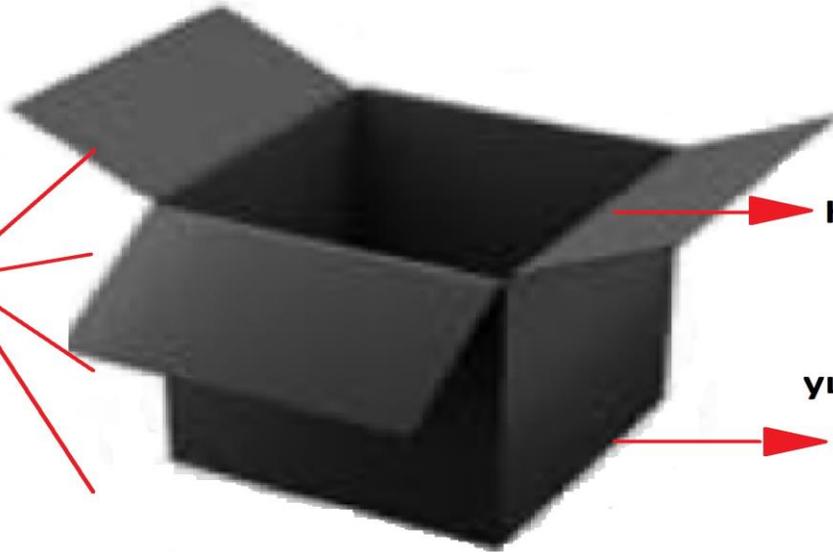


ПОЛИТЕХ

# МОДЕЛЬ С ТОЧКИ ЗРЕНИЯ ПОЛЬЗОВАТЕЛЯ С ОБЪЯСНЕНИЕМ



2 М пикселей



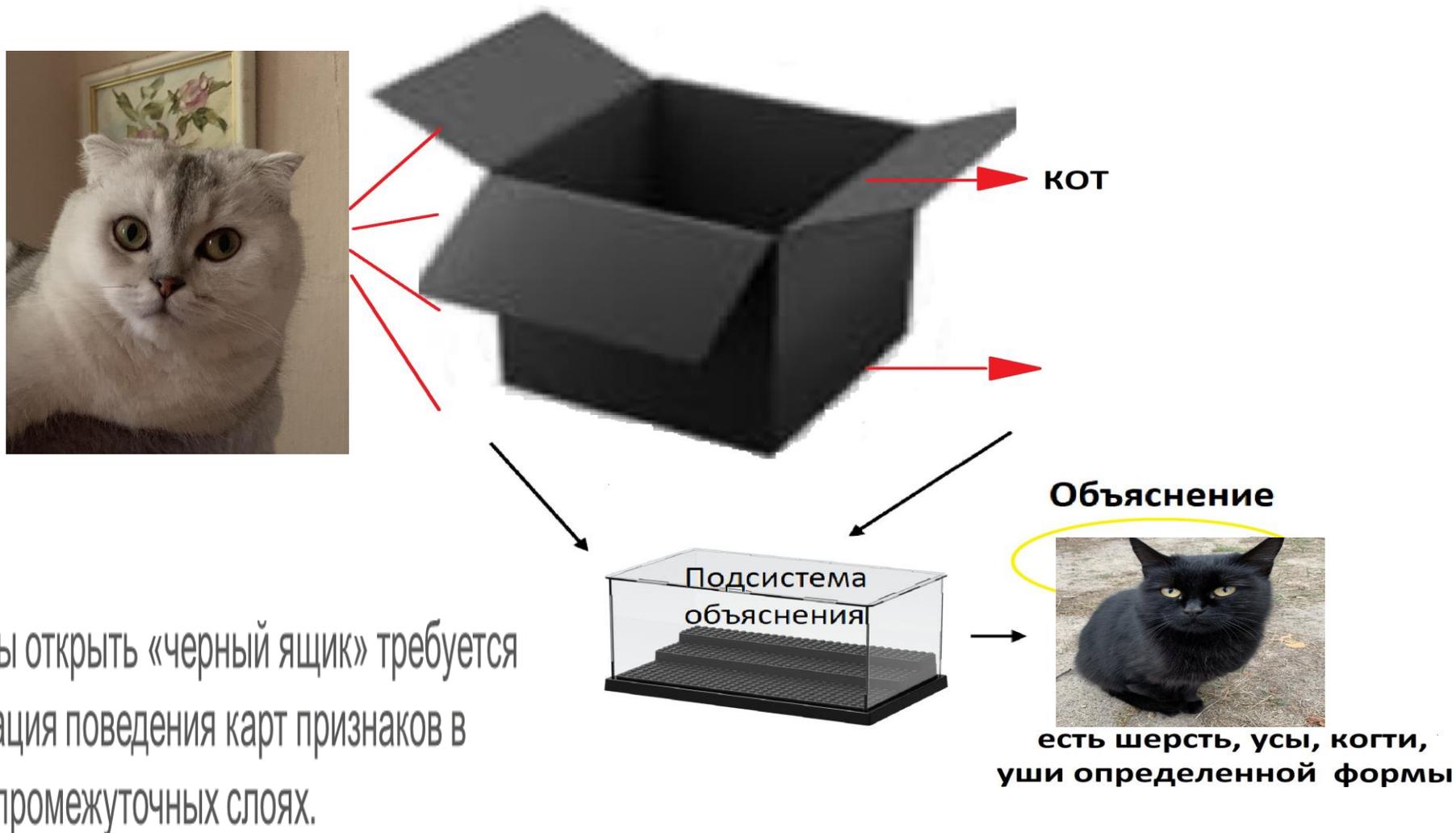
КОТ

есть шерсть, усы, когти,  
уши определенной формы

N понятий

«черный ящик» предсказывает, что на фотографии - **КОТ** с вероятностью 0.98, так как у него есть шерсть, усы, уши определенной формы»

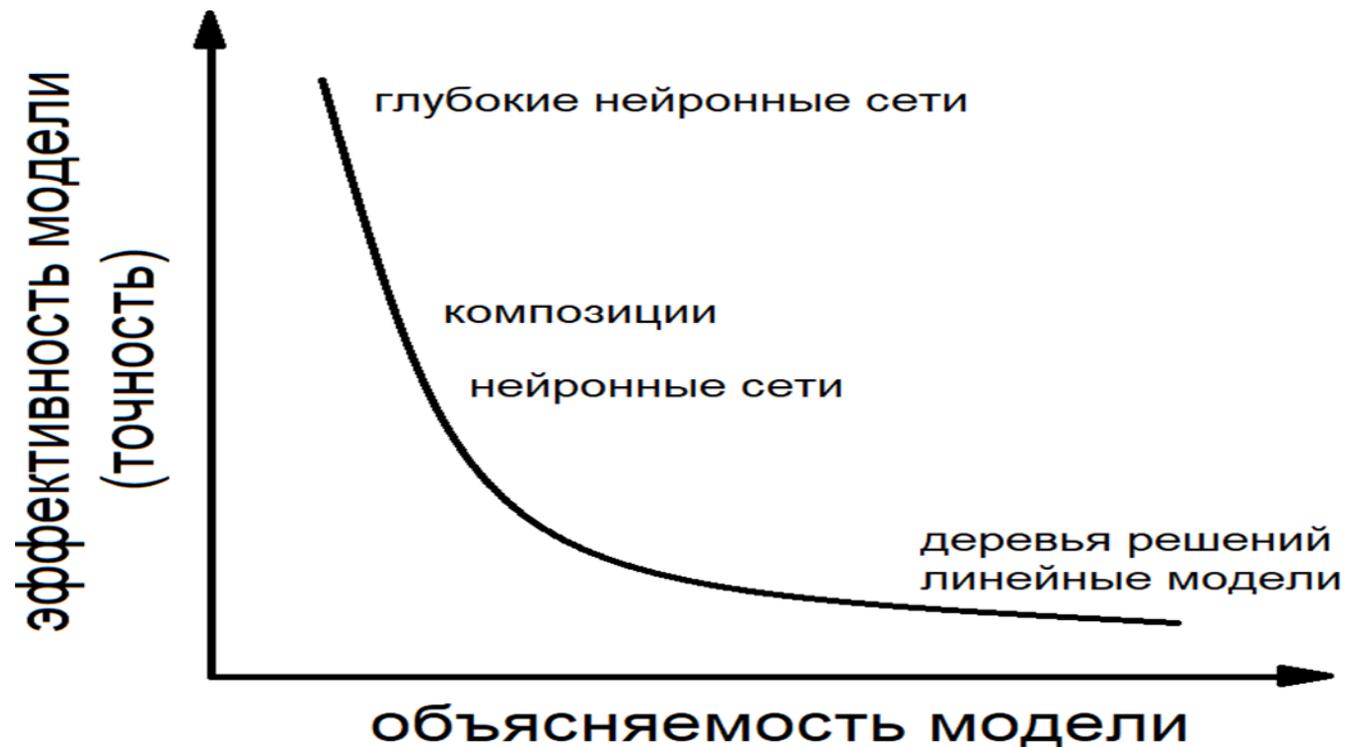
Должна быть не только система расчетов, но и подсистема дедуктивных объяснений в терминах понятий, объем которых N многократно меньше 2M



Для того, чтобы открыть «черный ящик» требуется интерпретация поведения карт признаков в промежуточных слоях.



# А НУЖНО ЛИ ОТКРЫВАТЬ «ЧЕРНЫЙ ЯЩИК»? ?



Да, нужно, так как **эффективные** (точные) модели «черного ящика» обычно сложны для понимания и не **самообъясняемые**, что снижает **уровень доверия к результатам**

- **программирования** МТ– следование принципу хранимой программы /данных и **последовательного** выполнения **статического** множества операций ..... булевой алгебры.
- машинное обучение искусственных нейронных сетей – «запутывание» программы /данных и **параллельное** выполнения операций отображения **динамического** множеств входных данных на множество выходных данных (аналог процессов с обострением)

:

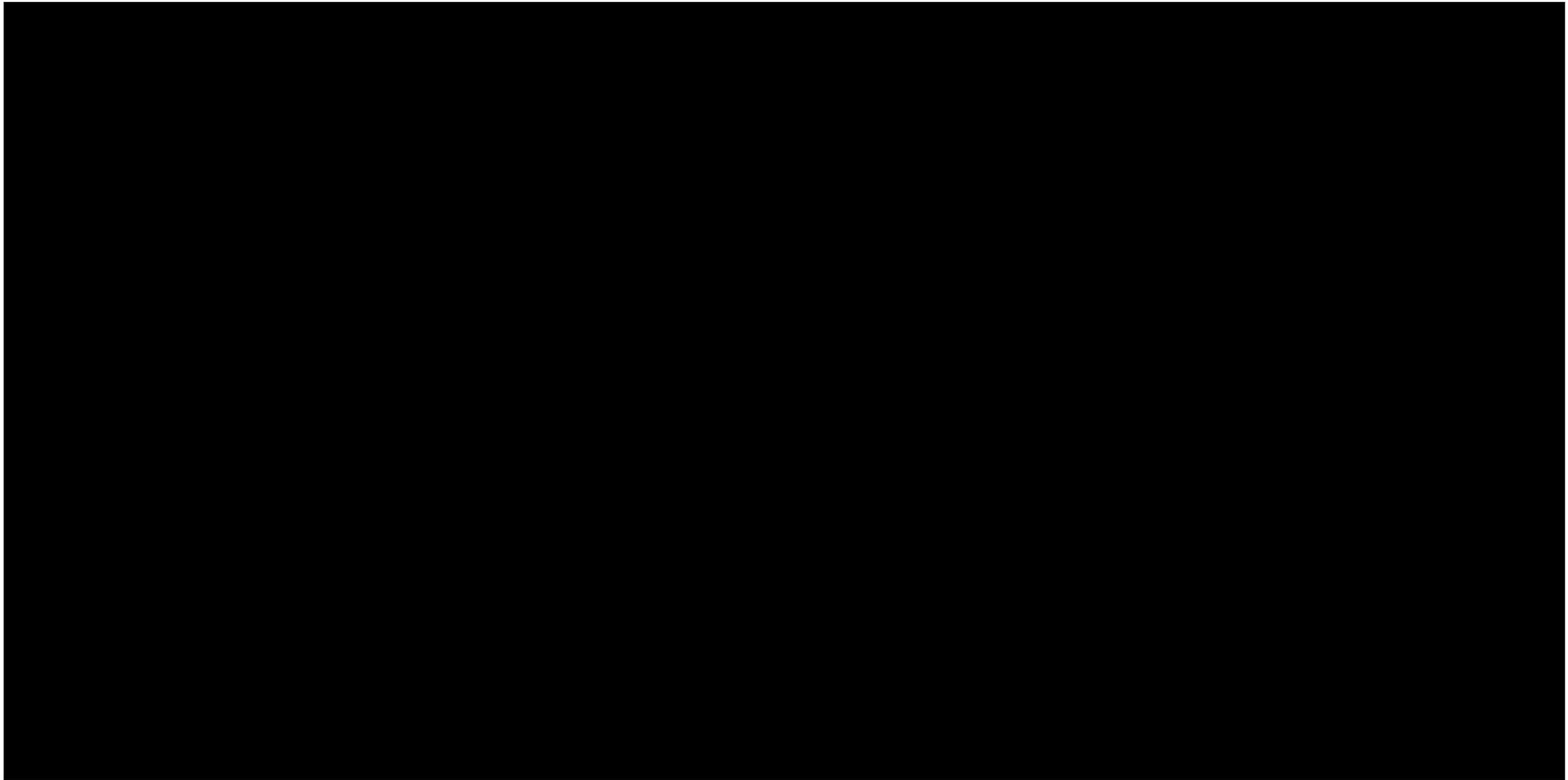
**Актуальность** этих двух подходов высокая, а повысить «остроту» компьютерных инструментов можно за счет «обучения» компьютерных систем реализовывать различные классы алгоритмов, реконфигурируя свою «вычислительную структуру».

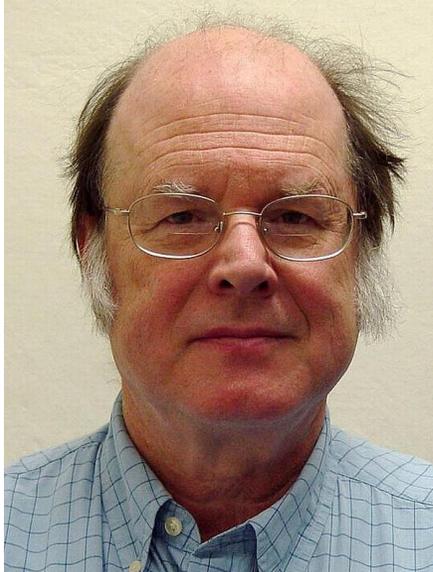
## Машина Тьюринга (MT)

## Нейронные сети (НС)

Процессор	Сложный	Простой
	Высокоскоростной	Низкоскоростной
	Один или несколько	Большое количество
Память	Отделена от процессора Локализована Адресация по расположению	Интегрирована в процессор Адресация по содержанию
Вычисления	<b>Централизованные</b>	<b>Распределенные</b>
	<b>Последовательные</b>	<b>Параллельные</b>
	<b>Хранимые программы</b>	<b>Способность к обучению</b>
Надежность	Высокая уязвимость	Высокая робастность
Специализация	Численные и символьные операции	Задачи восприятия
Среда функционирования	Строго определенная и ограниченная	Не определенная и без ограничений

Заметим, что связи между нейронами н/с играют роль "памяти» для хранения «программы» и настраиваются в процессе обучения или функционирования.





*The digital computer may be thought as an engine that dissipates energy in order to perform mathematical work. Цифровой (компьютер можно представить как «двигатель», который рассеивает энергию ,выполняя математическую работу)*

*Charles H . Bennett, 1981*

*Ч. Беннет*

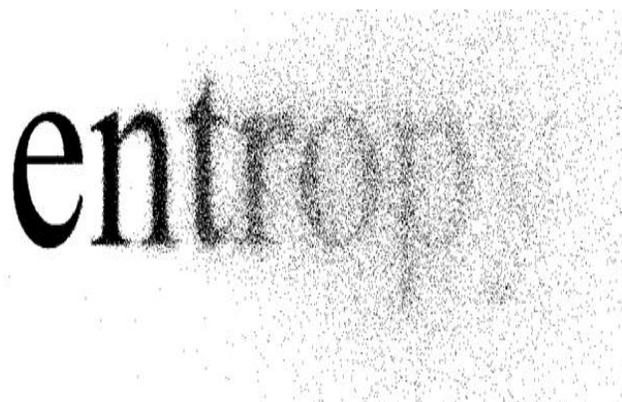
**Вопрос: в чем суть «математической работы» которая производится в ИНС с точки зрения физики?**

Формально любой алгоритм вычислений с точки зрения физики задает «траекторию» движения (чего) в «лабиринте» состояний конфигурационного пространства компьютера. Алгоритмом может быть «программа», но может быть сама структура «вычислителя»

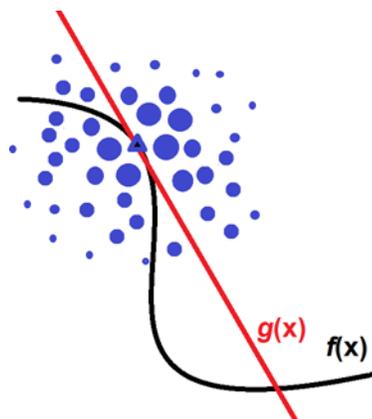
Можно ли так сформулировать *концепцию* «вычислений», так чтобы достигать «математически значимого» результата , при этом не только **потребляя энергию, но и объясняя полученные результаты ?**

Суть интеллектуализации компьютерных технологий **«нового поколения»**:

1) новая модель память-процессор, способная накапливать информацию о произведенных вычисленных и строить на их основе «суррогатную» модель для объяснения результатов решения задач и различных вариантов их используя ...



Неопределенность  
данных

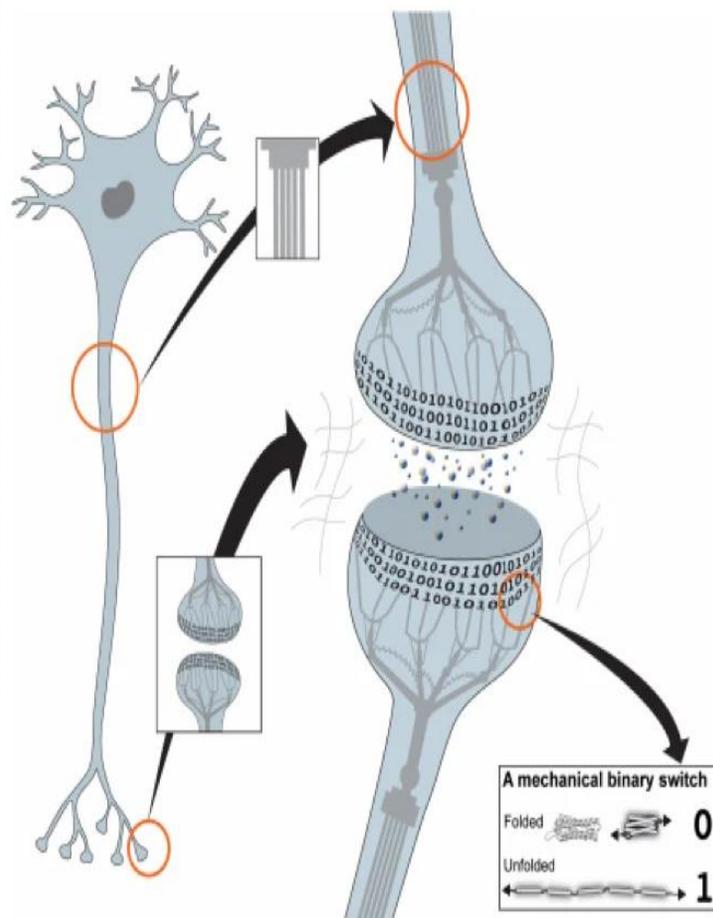
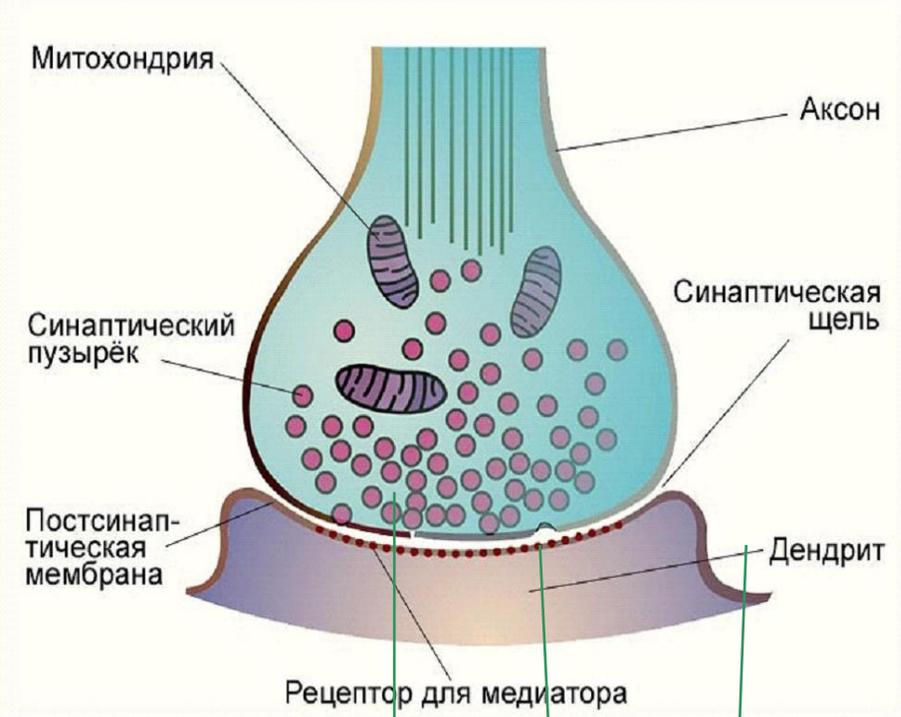


Модели  
объяснения

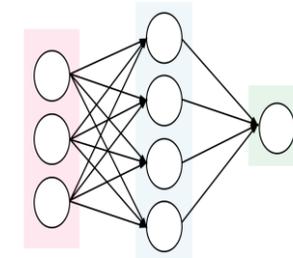
2) новый метод регуляризации с помощью использования «суррогатной» модели, используемой для объяснения того, почему полученное решение **«обратной задачи»** отвечает принятой «модели реальности»

Умение решать обратные задачи – основа интеллекта: в новых условиях «поиск» решений в пространстве возможностей организует и **объясняет сам интеллектуальный вычислитель.**

# ПРИМЕР РЕШЕНИЯ : БИОИНСПИРИРОВАННАЯ СРЕДА ВИРТУАЛЬНОГО МОДЕЛИРОВАНИЯ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ, ПРОТЕКАЮЩИХ В АССОЦИАТИВНОЙ ПАМЯТИ



Физические каналы передачи информации – цепочка аксонов, инициирующие нейромедиаторные **виртуальные каналы**, являющиеся частью пространства состояний для хранения «паттернов» - аналога **аддитивного ситуационного базиса**



Системная конфигурация «виртуальные» информационные каналы передачи нейромедиаторов (категорная конструкция)

Нейромедиаторы —целостный носитель (сборка) **данных и программ, представляемый** через множество морфизмов .

«Если не знаешь, к какой пристани  
плыть, ни один ветер не будет  
попутным».

*Сенека*



В нейронных сетях как пространственно-распределенных системах потенциально могут существовать **режимы с обострением** – процессы, которые можно рассматривать с позиций сгущения точек постоянной фазы, сохраняющих структуру паттерна

Суть научения - настройка **вычислителя** решать задачи исходя из требований

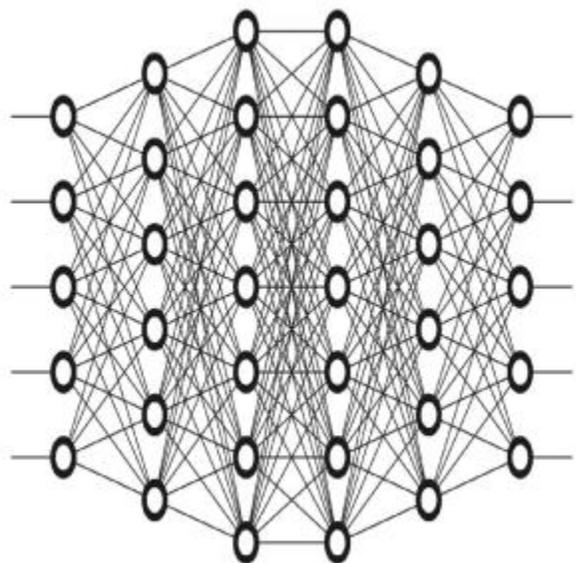
.... **Минимальной** диссипации энергии – .....

**Минимального** числа операций (транзакций процессор-память) доставляющих решение задачи

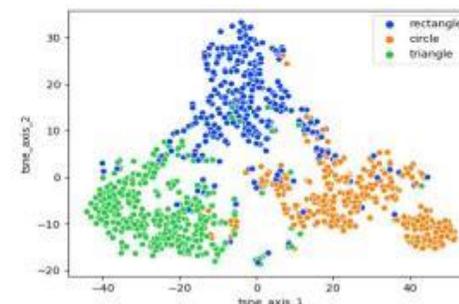
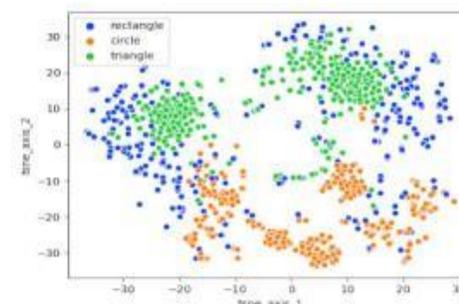
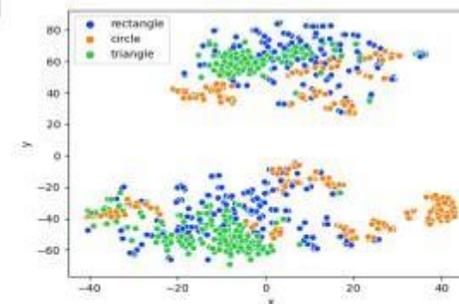
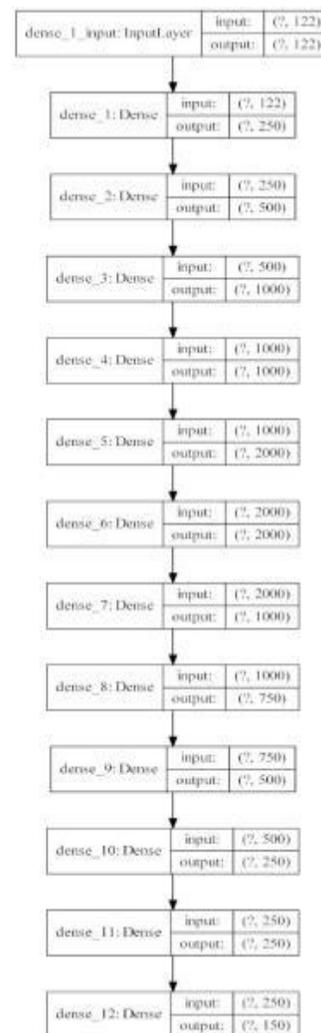
... **Объяснения** полученных результатов на основе конечной совокупности паттернов

Можно предположить, что именно в нейроморфной среде возможно возникновение **«процессов с обострением»** такого преобразования входных данных, при котором размерность пространства данных сокращается, но не искажаются паттерны, используемые в системе «интерпретации» полученных результатов.

# ПОИСК РЕШЕНИЙ В «ПРОСТРАНСТВЕ ВОЗМОЖНОСТЕЙ» НЕЙРОННОЙ СЕТИ

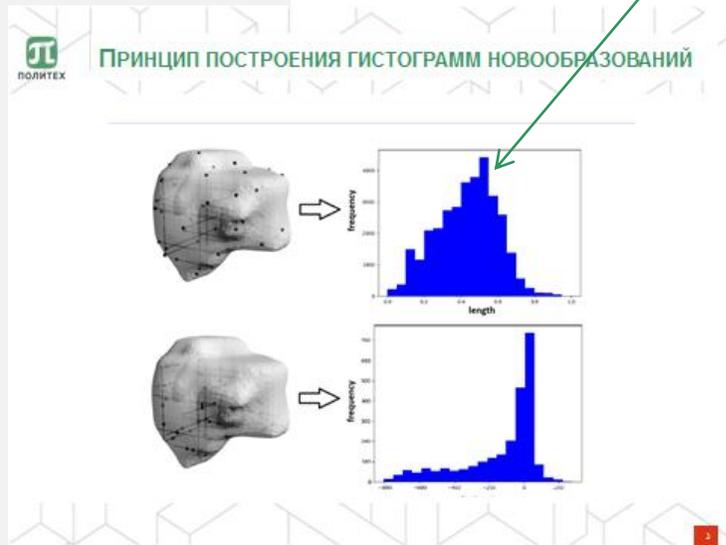


Завершения процесса обучения оцениваться на основе анализа структуры **аттрактора вектора состояний** слоев нейронной сети

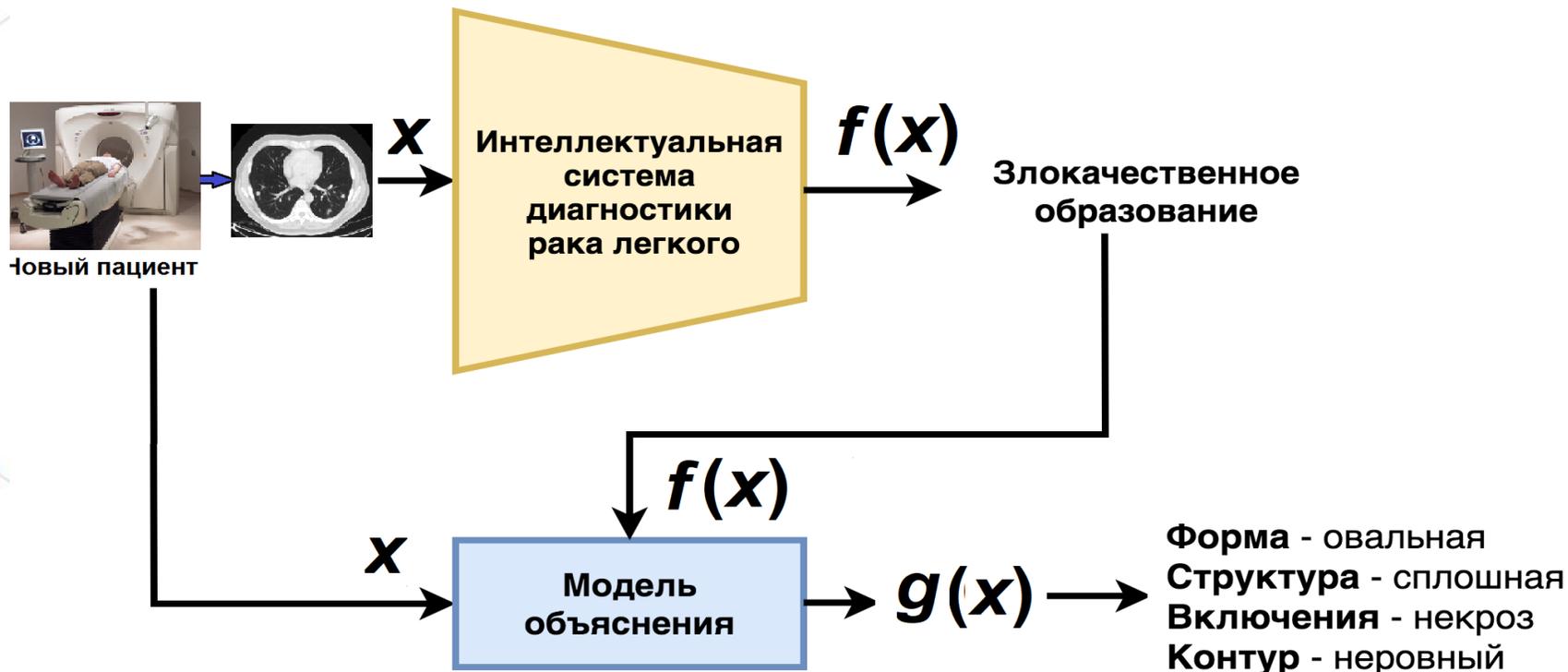


# ПРИМЕР: КАК РЕАЛИЗОВАТЬ НАУЧЕНИЕ НА ПРИМЕРЕ МЕДИЦИНСКОЙ ПРОБЛЕМАТИКИ

Как построить метод объяснения на основе объективных (не зависящих от данных) **инвариантов-патернов** (например, топологических) и модели МО (глубокая нейронная сеть, случайный лес, SVM и т.д.), которая «аппроксимирует» свойства сложной модели реальности **в окрестности конкретного примера**

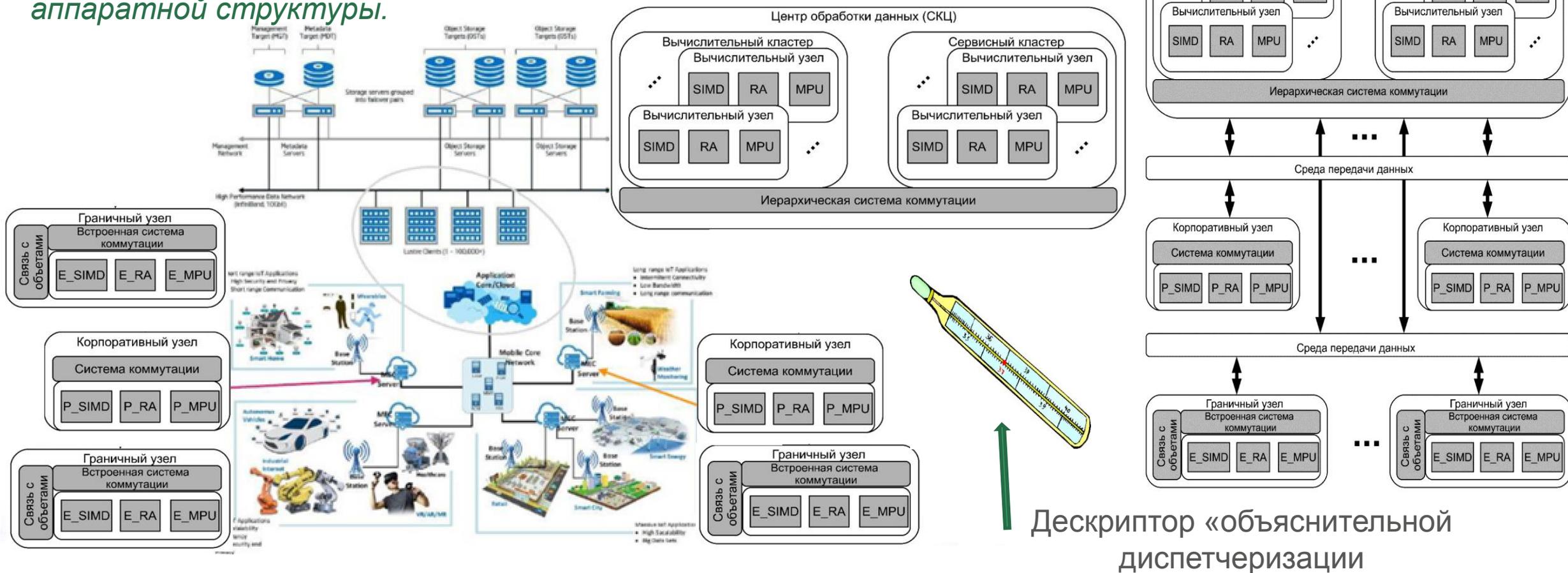


$$g(x) \approx f(x) \text{ в точке } x$$



# ПРИМЕР: СТРУКТУРА «УМНОЙ» ГЕТЕРОГЕННОЙ СИСТЕМЫ СКЦ «ПОЛИТЕХНИЧЕСКИЙ» С ИНТЕЛЛЕКТУАЛЬНЫМ ДИСПЕТЧЕРОМ

На компонентном и архитектурном уровнях в автоматическом режиме вычислитель адаптируется (обучение с подкреплением) к прикладному алгоритму путем реконфигурации предоставляемой аппаратной структуры.



- Интерпретируемость модели важна не меньше, чем числовая точность результатов расчетов.
- Совместное использование **суперкомпьютерных технологий и машинного обучения позволяет** выявить значение каждого их параметров программы моделирования путем интерпретации полученных результатов как с точки зрения различных числовых «мер» (площадь, длина, вес, скорость, ...), так и понятий.



- **Линейная регрессия**
- **Логистическая регрессия**
- **GLM (Лассо, гребневая регрессия)**
- **GAM (Обобщенная аддитивная модель)**
- **Деревья решений**
- **К ближайших соседей**



"LOOK DEEP INTO  
NATURE, AND THEN YOU  
WILL UNDERSTAND  
BETTER."

*ALBERT EINSTEIN*

Добро пожаловать в «суперкомпьютерное море» объяснительного интеллекта