

ВСИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

курс: Введение в профессиональную
деятельность

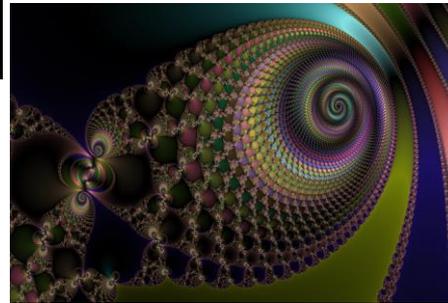
ЛЕКЦИЯ 9: ОСНОВНОЙ ВОПРОС «ЦИФРОВОЙ
ТРАНСФОРМАЦИИ»: ВОЗМОЖНОСТИ МОДЕЛИРОВАНИЯ
ИНТЕЛЛЕКТА С ПОМОЩЬЮ КОМПЬЮТЕРНЫХ ВЫЧИСЛЕНИЙ

6.04.2023



Вычисления функций

процессы мышления



Рекурсия - есть метод описания (объяснения) ситуации, когда объект (процесс) является частью самого себя:

- Либо во **времени**
- Либо в **пространстве**
- Либо и **там и там**

Механизм рекурсии лежит в основе иерархической организации вычислений и структуре всех объектов Природы:

- от наименьших наблюдаемых элементарных частиц до
- живых организмов и видимых скоплений галактик



Порождающие структуры объектов



ТАКЖЕ: ВВЕЛИ ПОНЯТИЕ ИНТЕЛЛЕКТУАЛЬНОЙ РЕКУРСИИ – ПРОЦЕССА СВЯЗЫВАНИЯ ПОТОКА ОЩУЩЕНИЙ В ПОТОК ПОНЯТИЙ

quod sentimus loquamur

мы **чувствуем**, что **говорим**

quod loquimur sentiamus

мы **говорим**, то что **чувствуем**

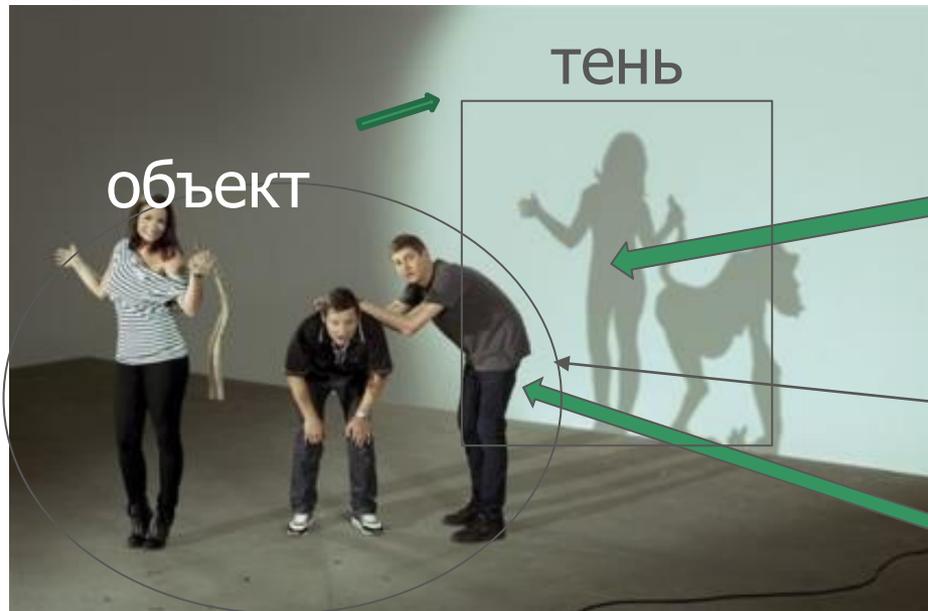
- Основная идея «интеллектуальной рекурсии», которая положена в основу современного научного знания, заключается в том, что воспринимаемые/доступные органам чувств данные «являются носителями как
 - **количественной меры и формы** – in_forma_tion,
 - так и
 - «смысла» – семантической меры.
- Интеллектуальная рекурсия позволяет **описать свойства** непрерывного разнообразия Природы, в которое погружены все существующие объекты, **конечным числом слов – носителей различных понятий.**
- Современная математика не более, чем формальный способ задания дискрипторов «различимосьи» данных наблюдений, но сам **способ формализации различий (коды)** может быть произвольный.

Итак. Компьютерные технологии лишь «кодируют» и «масштабируют» формализованные математикой различия



ПОЛИТЕХ

ПРОБЛЕМЫ ФОРМАЛИЗАЦИИ, ТЕОРЕМЫ ГЕДЕЛЯ - ТЕНИ ФИЗИЧЕСКОЙ РЕАЛЬНОСТИ



Символы (числа ?)

граница



Физические объекты (процессы)

Известна физическая **невычислимость** (необратимость) «прошлых» состояний термодинамических процессов, **однако** имеет место **полная обратимость** во времени и изотропность в пространстве **фундаментальных законов.....** (согласно т.1,2 Геделя такие законы либо противоречивы , либо неполны)

До сих пор нет ясности какие математические описания адекватны всем свойствам физической реальности.

Аксиома 1: истинно лишь то, что можно

- Вычислить на цифровом компьютере (? , «доказать»)
а полученный результат
- Понять и объяснить (? неформализуемость истны)

Аксиома 2. Точное решения «**прямых**» задач,
которые сводятся к результатам выполнения команд
программ-алгоритмов, можно получить с помощью
счетного числа **частично рекурсивных операций** ?

Аксиома 3. Объяснение решения «**обратных**» задач
или интерпретации результатов не вычислимая
функция

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»

ИНТЕЛЛЕКТУАЛЬНЫЙ СИНТЕЗ: КОНСТРУИРОВАНИЕ «НОУМЕНА» ИЗ КОМПОНЕНТ «ФЕНОМЕНА»

ПОЛИТЕХ





ПОЛИТЕХ

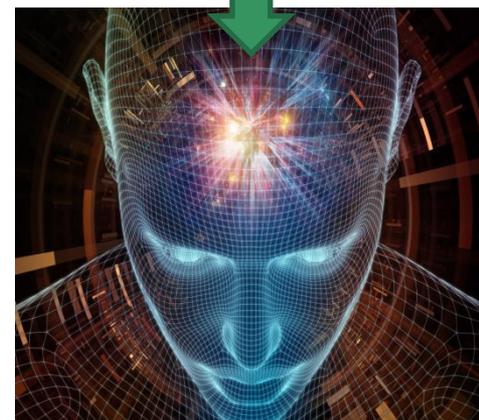
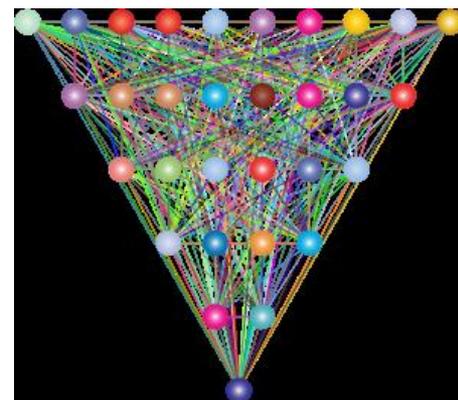
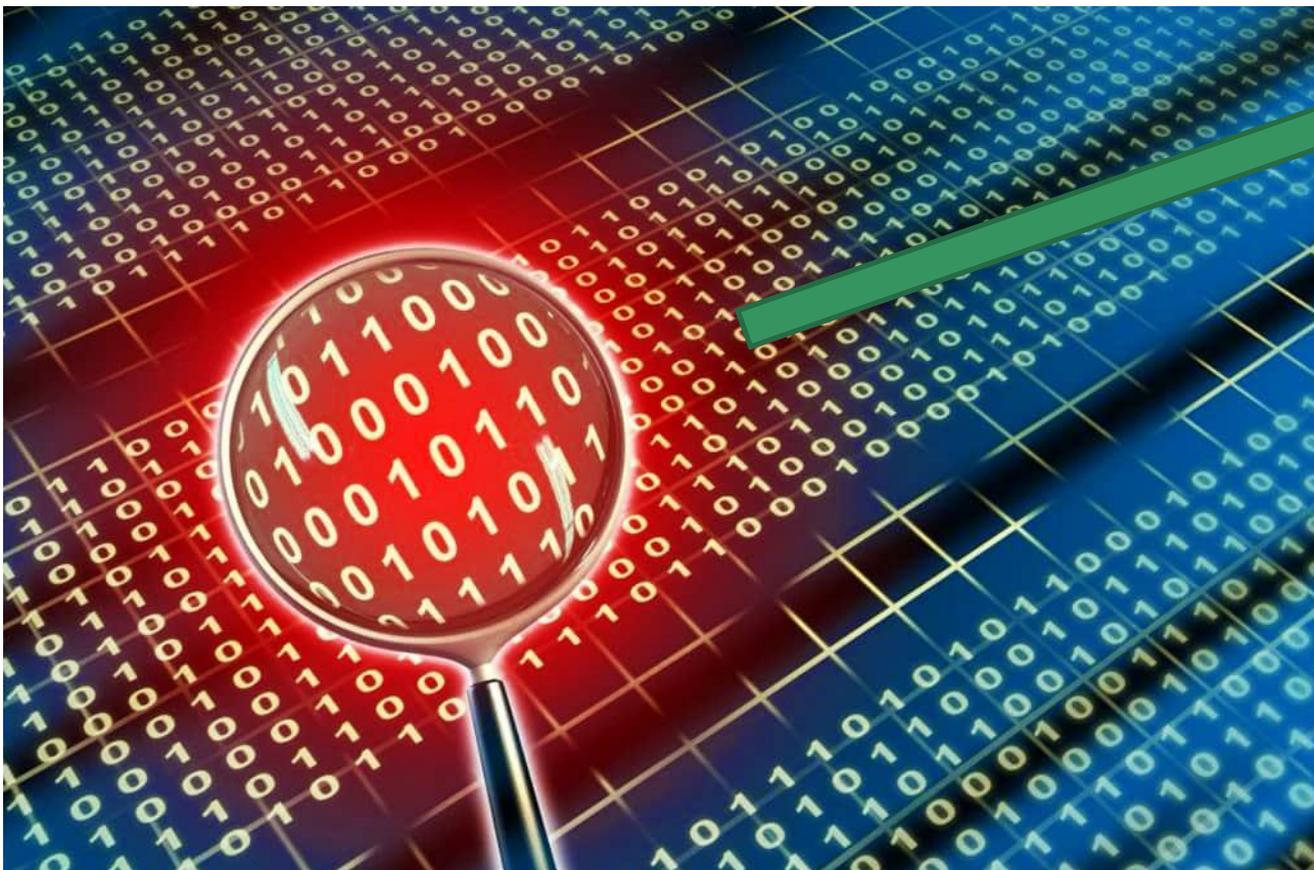
ПРОБЛЕМА ФОРМИРОВАНИЯ «ЭКОСИСТЕМЫ ЗНАНИЙ»

Что же требуется чтобы понять и объяснить окружающий мир?

Ответ: «необходимо умение создавать и понимать ... то, чего в **целостном** виде еще нет, используя для этого все «большие данные», предыдущий опыт и теоретические знания».

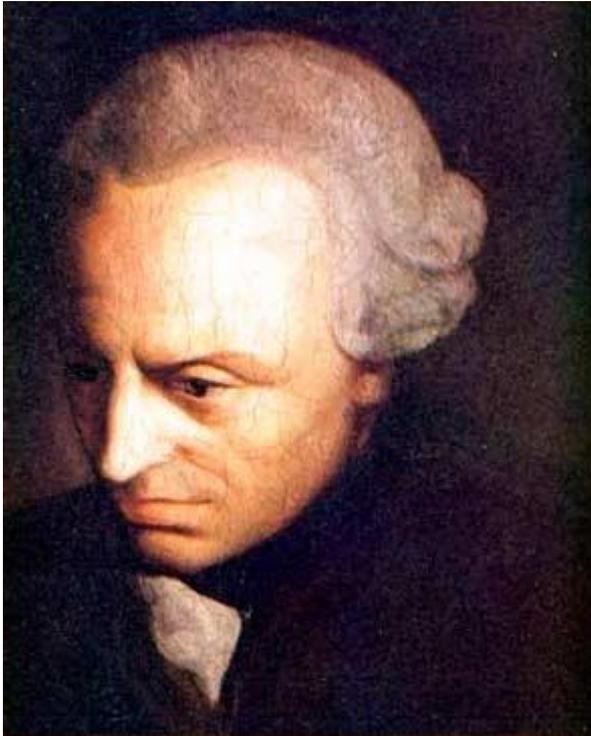
Что для этого необходимо: **целостное** восприятие данных, порождаемых объектами реального мира в контексте метафоры «**знание – сила**».





Цифровой код – однородное и делимое, ...
восприимчивое к частностям, случайности и
основанное на понятие «**равенства количеств**»

Когнитивное целостное –
результат познания,
основано на понятии



Ключевой вопрос - как «далеко» **человеческий разум** ... или его аналог - «искусственный интеллект» , используя опыта, носителем которого являются цифровые данные (т.н. феномены), **может пройти в познании мира** и ... управлении предметами и явлениями.

Суть философии «критики» И. Канта - рациональная идея **познания** вне опыта ограничена, хотя только локального текущего опыта без обобщений для **познания** не достаточно.

Если за меру сложности вычислений взять «число степеней свободы» объекта в пространстве состояния решаемой задачи, то за «единичную меру» сложности можно взять 1D (одномерную) задачу, аргумент которой разбит на 100 частей, т.е.:

$$1^{100} = \langle 1 \rangle$$

Если в задаче 1D «интервал» заменить на «квадрат» 2D , то есть увеличить размерность задача в 2 раза, то мера «сложности» станет

$$2^{100} = 1,2676506 \times 10^{30} \approx \langle 1 \rangle * \infty$$

Вывод: Вычислительная сложность «реальных» задач науки и техники требует использования специальных инструментов – суперкомпьютеров, в которых объем памяти и быстродействие процессоров позволяет решать задачи с заданной точностью **за приемлемое или «конечное время»** .

СКОЛЬКО НУЖНО ЧИСЕЛ, ЧТОБЫ «ЗАКОДИРОВАТЬ» ВСЬ МИР ?

- Информация в компьютере записывается в виде «нулей и единиц»
Так, чтобы сохранить фотографию объемом 3 мегабайта надо $3 \times 8 \times 2^{20}$ т.е. больше 25 миллионов ячеек памяти компьютера
- Данные для обработки могут храниться в оперативной памяти или на «жестком диске». Объем хранимых данных практически ни чем не ограничен, но ограничена скорость доступа к хранимым данным.
- Метод «LogLog обработки» - преобразование массива входных данных произвольной длины в (выходную) битовую строку фиксированной длин:
 - $\log_2(\log_2(1\,000\,000\,000)) = 4.9$ Каждой записи присваивается т.н. хэш-значение, т.е. вычисляется свертка. Возвращаемые хеш-функцией значения (выходные данные) менее разнообразны, чем значения входного массива (входные данные). Пример - хеш-функция остаток от деления входных данных на M . Не всегда, если записи совпадают, то и их хеш-значения тоже совпадают.



Предельные возможности современных «компьютеров»

Так, минимально возможные линейные размеры транзистора (x_{\min}) и максимальная рабочая частота ($f_{\max} = 1/t_{\min}$) ограничиваются соотношения неопределенностей Гейзенберга (h – постоянная Планка):

$$\Delta x * \Delta p = x_{\min} * \Delta p \geq h/2\pi, \quad \text{где } \Delta p = \sqrt{2} * m * E_{\text{bit}} = \sqrt{2} * m * kT * \ln 2$$

$$\Delta t * \Delta E = t_{\min} * \Delta E \geq h/2\pi, \quad \text{где } \Delta E = kT * \ln 2$$

при $T=300$ К, получим $x_{\min} = 1.5$ нм, а $t_{\min} = 4 * 10^{-14}$ с .

Если создать микропроцессор, у которого одновременно будет и самая большая плотность упаковки (определяемая x_{\min}) и максимально возможная частота (определяемая t_{\min}), и оценить, какая при таких условиях должна выделяться мощность P на единицу площади, то при $T=300$ К, получим:

$$P = E/t = n_{\max} E_{\text{bit}}/t_{\min} = kT * \ln 2 / ((x_{\min})^2 * t_{\min}) = 3.2 * 10^6 \text{ Вт/см}^2,$$

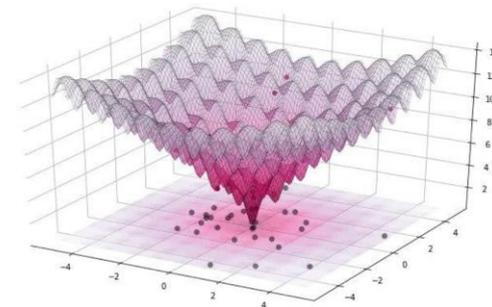
где n_{\max} – число транзисторов, которые занимают площадь, определяемую линейными размерами канала (x_{\min}).

Логическая схема редукционизма (механики Ньютона) проста – целое состоит из частей. А для того, чтобы **понять все целое – достаточно понять все его части.**

Но в 20 веке было открыто свойство «эмерджентности» : **целое «реальнее» своих частей:**

- дифракция электронов показала, что **квантовые частицы** не имеют определенных траекторий (**принцип неопределенности** Гейзенберга);
- центральным понятием квантовой механики – является не материальная частица, а ее т.н. «волновая» пси-функция, которая принципиально **не может быть измерена никаким прибором**, то есть **является невещественной сущностью**, в которой материи как таковой нет.
- Только умозрительное, т.е. доступное **исключительно разуму человека**, позволяет понять свойства того, что **чувственно воспринимается или измеряется приборами.**
- пси-функция **системы** всегда адекватнее **описывает ее свойства**, чем совокупность пси-функций, относящихся к ее **частям**, взятым по отдельности

- Перечислимость множеств,
 - Вычислимость функций,
 - разрешимость множеств...



- **Объяснимость (интерпретируемость) результатов**

Имеется потребность использовать мультимодальные метрики, которые позволяют определять какие:

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



Новые понятия: объяснение, интерпретация, понимание, доверие

От результатов вычислений требуются

- 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.

нет трудных задач, а есть задачи, которые
трудны для определенного уровня понимания
того, как **использовать** полученные результаты

.....

- **Интеллектуальные вычисления** способны манипулировать числами как абстрактными символами и... объяснять полученные результаты («вычислять» слова)
- **"Входы" систем интеллектуальных вычислений** представляются в виде кодов из некоторых символов а выход может быть либо в виде абстрактного символа, либо в виде физического действия или обоих этих состояниях одновременно.
- **Вычислительная теория интеллекта** - это теория построения т.н. **«трансформеров»**, которая
 - описывает обратимые преобразования различных кодов из слов на естественном языке в числа или другие математические структуры (вектора, матрицы, тензоры...)
 - Обеспечивает процесс **уменьшения энтропии описания внешней среды посредством «вычисления» объяснения происходящих в этой среде процессов.**

Машина Тьюринга (МТ)

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	Централизованные	Распределенные
	Последовательные	Параллельные
	Хранимые программы	Способность к обучению
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.



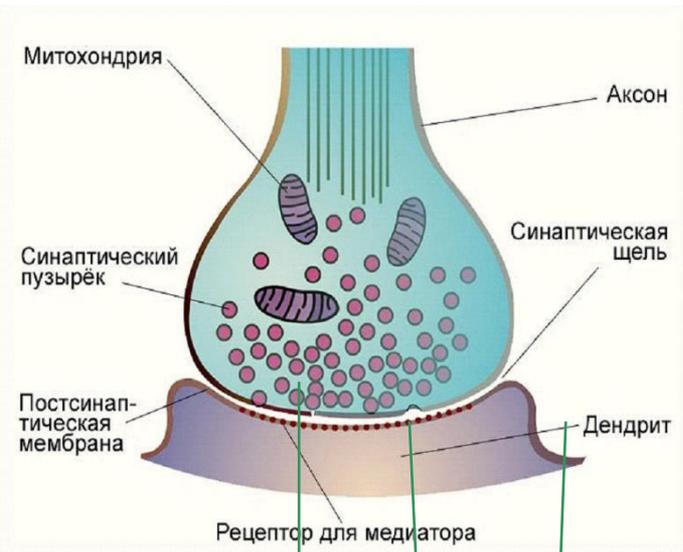
The digital computer may be thought as an engine that dissipates energy in order to perform mathematical work. Цифровой (компьютер можно представить как «двигатель», который рассеивает энергию, выполняя математическую работу)

*Charles H. Bennett, 1981
Ч. Беннет*

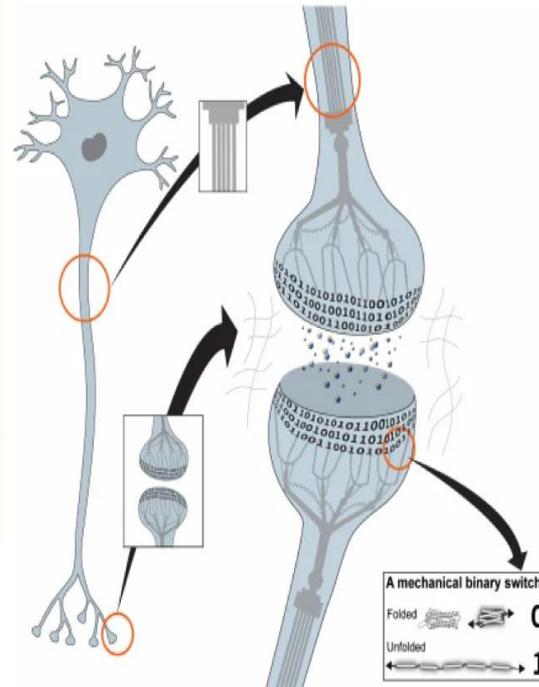
Вопрос: в чем суть «работы» которая производится «умными» компьютерами с точки зрения физики?

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 «формально значимого» результата, при этом не только **потребляя энергию, но и объясняя полученные результаты** ?.

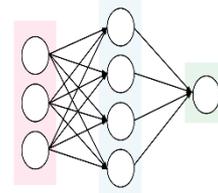


Системная конфигурация «виртуальные» информационные каналы передачи нейромедиаторов (категорная конструкция)

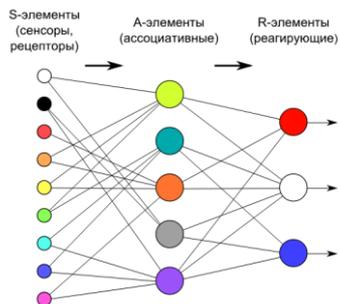


Физические каналы передачи информации – цепочка аксонов, инициирующих нейромедиаторные **виртуальные каналы**, являющиеся частью пространства состояний для хранения «паттернов» - аналога **аддитивного ситуационного базиса**

МОДЕЛЬ

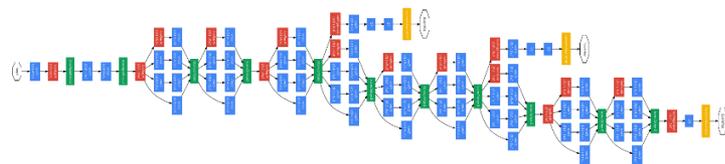


Нейромедиаторы —носители (сборка) **данных и программ, представляемый** через множество каналов .



Многоуровневый перцептрон

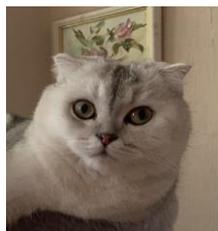
- классификация
- функциональная аппроксимация
- автокодер



Сверточные нейронные сети

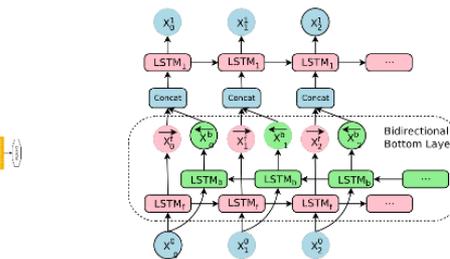
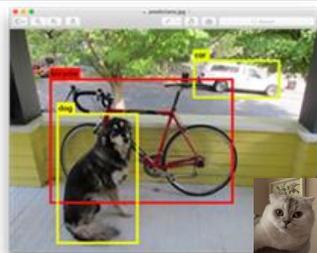
- выделение свойств
- детектирование объектов
- сегментация изображений

классификация



“КОТ”

Обнаружение объектов



Рекуррентные нейронные сети

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

Сегментация данных



Перцептрон - модель восприятия информации мозгом, состоящая из элементов трех разных типов: сенсоров, ассоциативных элементов и реагирующих элементов

Пример: задача обучения распознаванию 3D изображения компьютерной томографии (КТ)

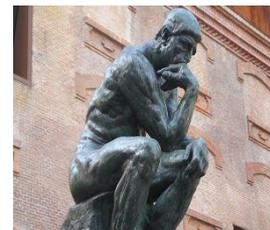
Обучающая выборка **10⁴ КТ снимков**



СК: **1 Пфлопс**
= **1 × 10¹⁵ Флопс**

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

КТ легких

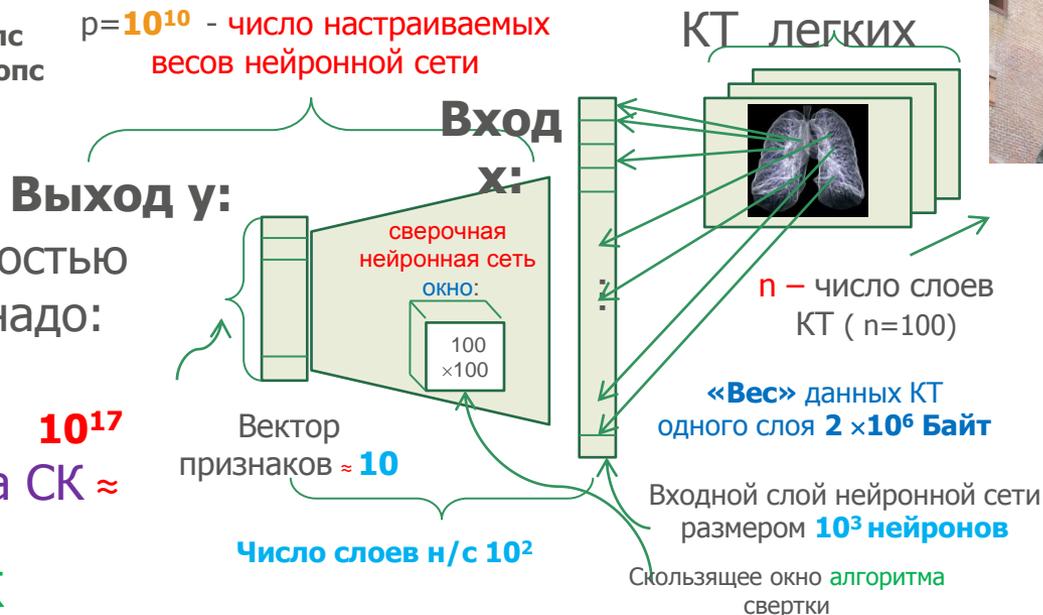


Для обучения н/с с точностью классификации **90%** надо:

- 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



обучение нейронной сети даже среднего размера может занимать **недели или месяцы**

Градиентный алгоритм обучения глубокой сверточной нейронной сети:

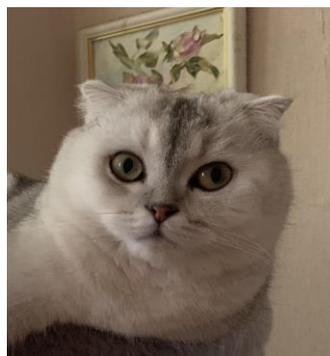
- функция ошибки $F = \|y^* - y\|^2$, y^* - эталонный вектор признаков
- **Алгоритм использует 10^{10} частных производных F** по всем настраиваемым параметрам;
- Число операций численного **дифференцирования** на одну итерацию **$Q = 10^{15}$**



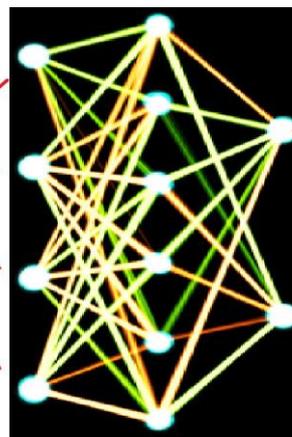
ПОЛИТЕХ

СТАНДАРТНАЯ МОДЕЛЬ МО С ТОЧКИ ЗРЕНИЯ РАЗРАБОТЧИКА

Контекст: кем,
чем, когда
сформировано
входное
множество ?



Входное
множество
– массив
из 2М
пикселей



операто
ρ

кот

пе
с

Выходное
множеств
о
из двух
понятий

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



Объем входных данных m

$$m > n$$

Объем выходных данных n

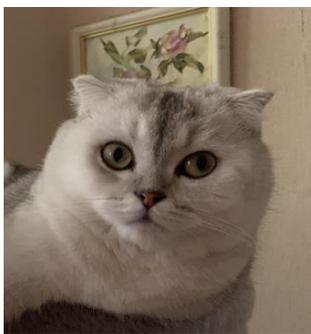
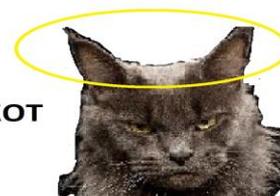


Фото 2 М пикселей



КОТ

есть шерсть, усы, когти,
уши определенной формы



пес

n понятий, используемых в объяснении полученного «решения»

«черный ящик» системы распознавания предсказывает, что на фотографии - **КОТ** с вероятностью 0.98,

так как у него есть шерсть, усы, уши определенной формы"

У «интеллектуальной системы» должна возможность не только выполнить расчеты, но и синтезировать набор дедуктивных заключений или предикатов, определенных в терминах понятий, объем которых n многократно меньше m



ПОЛИТЕХ

МОДЕЛЬ РАСПОЗНАВАНИЯ С ТОЧКИ ЗРЕНИЯ ПОЛЬЗОВАТЕЛЯ С УКАЗАНИЕМ КОНКРЕТНОГО ЭЛЕМЕНТА В ЗАДАННОМ КЛАССЕ

Задача:

Распознавание класса
и сопоставление
имени

Класс – «кот»
Имя «федя»



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

Механизмов интерпретации классов много, но у каждого объекта из данного класса может быть только одно имя



- Развитие искусственного интеллекта - фундаментальная задача компьютерных наук, которая расширяет зону «вычислимости» явлений Природы до границы решений, ПОЛУЧАЕМЫХ С ПОМОЩЬЮ цифровых технологий.
- Глобальное распространение систем ИИ повлияет на всю человеческую цивилизацию, также как и создание
 - паровоза, ДВС, ИСЗ,
 - микропроцессора, персонального компьютера,
 - Сети интернет
- Эти процессы несут **прямую угрозу** «эко системе человеческого интеллекта». Купирование угрозы требует внедрения новых подходов к обучению на основе **целостного восприятия** математических принципов кодирования (описания) явлений Природы, построения на базе этих кодов многообразий понятий и соответствующих алгебраических систем, используя которые можно **прогнозировать последствия** и **объяснять «вычисляемые» с их помощью решения.**