

ВСИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

курс: Введение в профессиональную деятельность

ТЕМА 4 . ИДЕИ И ТЕХНОЛОГИИ ИИ

ЛЕКЦИЯ 14 РЕШЕНИЯ «ОБРАТНЫХ ЗАДАЧ» И МЕТОДЫ
ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ.

15 мая
2025

ПОЯВЛЕНИЕ ЦИФРОВОГО ИНТЕЛЛЕКТА - ПРИНЦИПИАЛЬНЫЙ ВЫЗОВ ПАРАДИГМЕ СОВРЕМЕННОЙ НАУКИ

- Биологические системы развивались миллионы лет, приспосабливаясь к различным условиям и демонстрируя когнитивные способности решения ограниченного круга «прикладных задач» своего выживания
- Изучение процессов функционирования биологического интеллекта позволило создать весьма «простые» нейро-подобные модели, позволяющие свести алгоритмы работы систем «искусственного интеллекта» (ИИ) к формальным математическим методам, но не поняв до конца механизм работы «естественного интеллекта»

Появление «лингвистически ориентированных» компьютерных систем стимулирует :

- 1) разработки новых моделей описания, накопления и передачи знаний,
- 2) создание компьютерных экзо-интеллектуальных систем использования «цифровых знаний» в контексте не формализуемых целей, реализуя при этом ряд преимуществ :

- передачи и сохранить в цифровом виде знаний, реализующих супервентность наблюдаемых явлений (набор физических свойств) и мыслимых понятий (тезаурусом)
- возможность работать с огромными объемами данных, обрабатывать информацию в реальном масштабе времени]6 реализуя процессы обучения

Наилучшим «модельным прототипом» открытых экзо-интеллектуальных систем является «мультипротокольная сеть «Интернет», дополняющая естественный интеллект механизмами «генеративного поиска» по запросам, а-ля GPT

Существует различие между числом, словами и понятиями.

В основе этого различия лежит способность «феномена мышления» **объединить формальные и экспериментальные моделей описания реальности ...**

Для этого можно использовать различные методы, которые базируются на интерпретации реальности с использованием таких понятий как

- квалиа
- супервентность
- смысл
- мыслимое
- величина
- и пр.

Когда верно «символическое» равенство ?

$$0=1$$

Что такое «единица» и «что такое» «ноль» ???

ТАК, ПОНЯТИЕ «МЫСЛИМОЕ» ЭТО СИМВОЛЬНЫЙ «КОД» СУПЕРВЕНТНЫЙ ФИЗИЧЕСКОМУ «ПРОЦЕССУ»

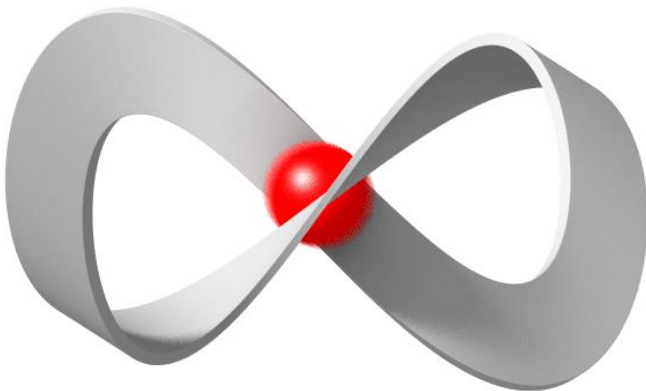
«материальное» можно определить через «обратное» как то, что отличается от информационного.



Музыка – это – ноты или звуки ... все зависит от контекста

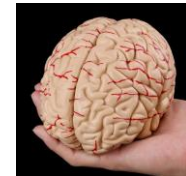
Гипотеза Джеффри Хинтона (автор метода «глубокого обучения», лауреат Нобелевской премии по физике 2024):

- «**Биологический интеллект** нужен был Природе, чтобы появился **«цифровой интеллект» ???**
 - Мотивация: цифровой интеллект требует для своего повеления (развития) много энергии, поэтому нужен был "биологический интеллект", создавший современную



*«My view is throw it (AI)
all away and start again»*

Джеффри Хинтон



Многие нуждаются не в Истине, а лишь в подтверждении своих догадок.....

- Переход физической материи в энергию $E=mc^2$ и наоборот — $Q \Rightarrow k \cdot T \ln 2$ выделение тепла при потере 1 бита информации (неравенство это удивительное свойство вещества Вселенной, позволяющее делать возможным любые материальные преобразования и воплощения, используя информацию.
- Энергия мысли воплощается в материю только если логически необходимые последствия этих воплощений **всегда соответствовали образам и свойствам реальных предметов**, то есть последствия образов были опять образами последствий с меньшим числом пустых (логически невозможных) отношений .

- Человек в процессе жизни чаще всего не достигает объективных собственных пределов
- Эти пределы формируются совокупностью знаний всего социума и доступными технологиями .
- Как понять, где проходит «граница» возможностей ИИ ?
- Может ИИ **достигнуть границ знаний**, сформированных всем социумом ?

«метафора: озноб узнавания»

- Античные греки —**принципиально избегали** использования бесконечности для объяснения своих взглядов («**бесконечность**» **не существует**) .
- Однако уже в средние века **строгость античных концепций отбросили** ради получения новых результатов.
- Хотя символ (∞) в современной науке широко используется в рамках обоснования доказательств, но при этом не входит в аксиоматику компьютерных наук.
- Введение «конструктивного» (а не только формального) смысла для понятия (∞) **открывает возможности решения различных «вычислительных» проблем** от «комбинаторного взрыва» до проблемы формирования обучающей выборки в задачах ИИ, решение которых «доказательно объяснимо с помощью конечного числа мыслимых понятий (суть проблемы **«квадратура цифрового круга»** в задачах ИИ)

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 ?

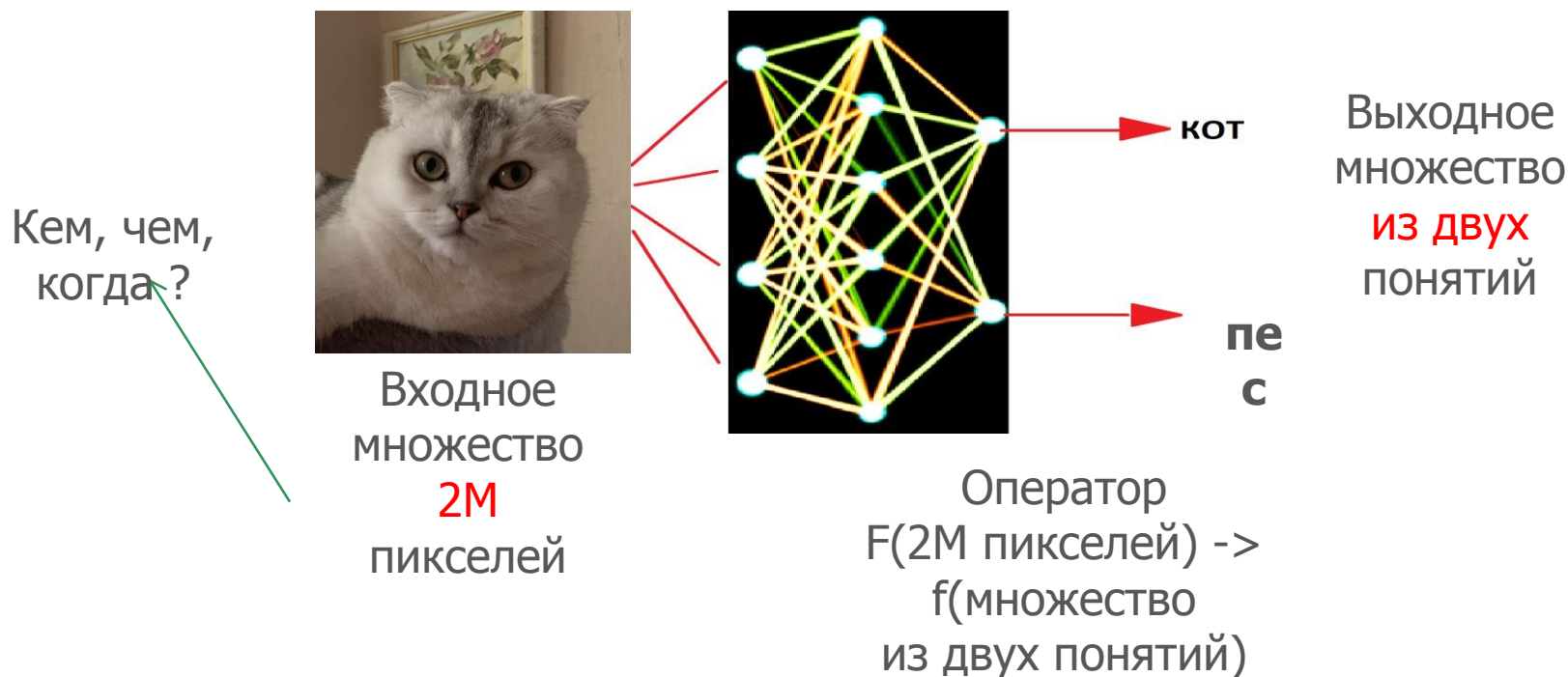
2) **Интеллектуальность** - решение «**обратных**» задач путем построения АЛГОРИТМОВ вычислений и объяснения (интерпретации) результатов

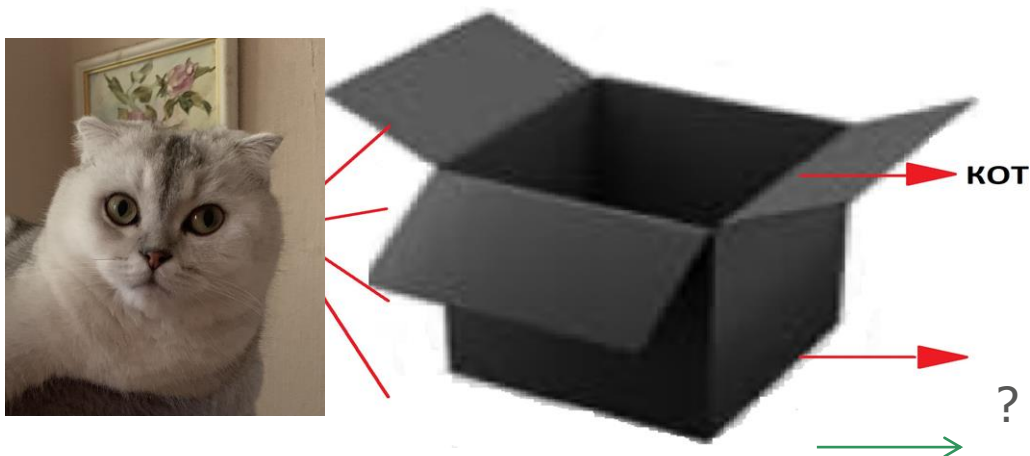
Интерпретируемость (объяснение) результатов - это степень уверенности ... в том, что вычисленный результат можно **понять**, а используемый алгоритм вычислений **уточнить** так, чтобы повысить степень уверенности и снизить «коэффициент незнания»

Ключевые слова: объяснение, интерпретация, понимание, доверие

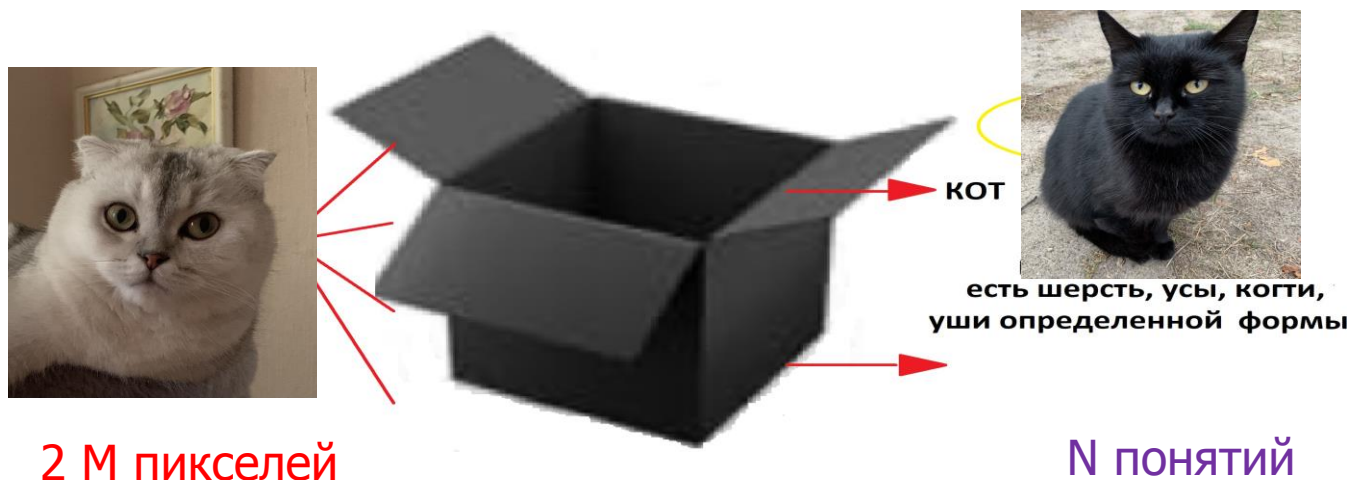
- Дедуктивные **объяснения** (explanation) в терминах понятий
- Индуктивная **интерпретация** (interpretation) в терминах параметров
- **Понимание** сделанных предсказаний (understanding) например с помощью SHAP(SHapley Additive exPlanation)
- **Доверие** к результатам функционирования модели (trust)

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





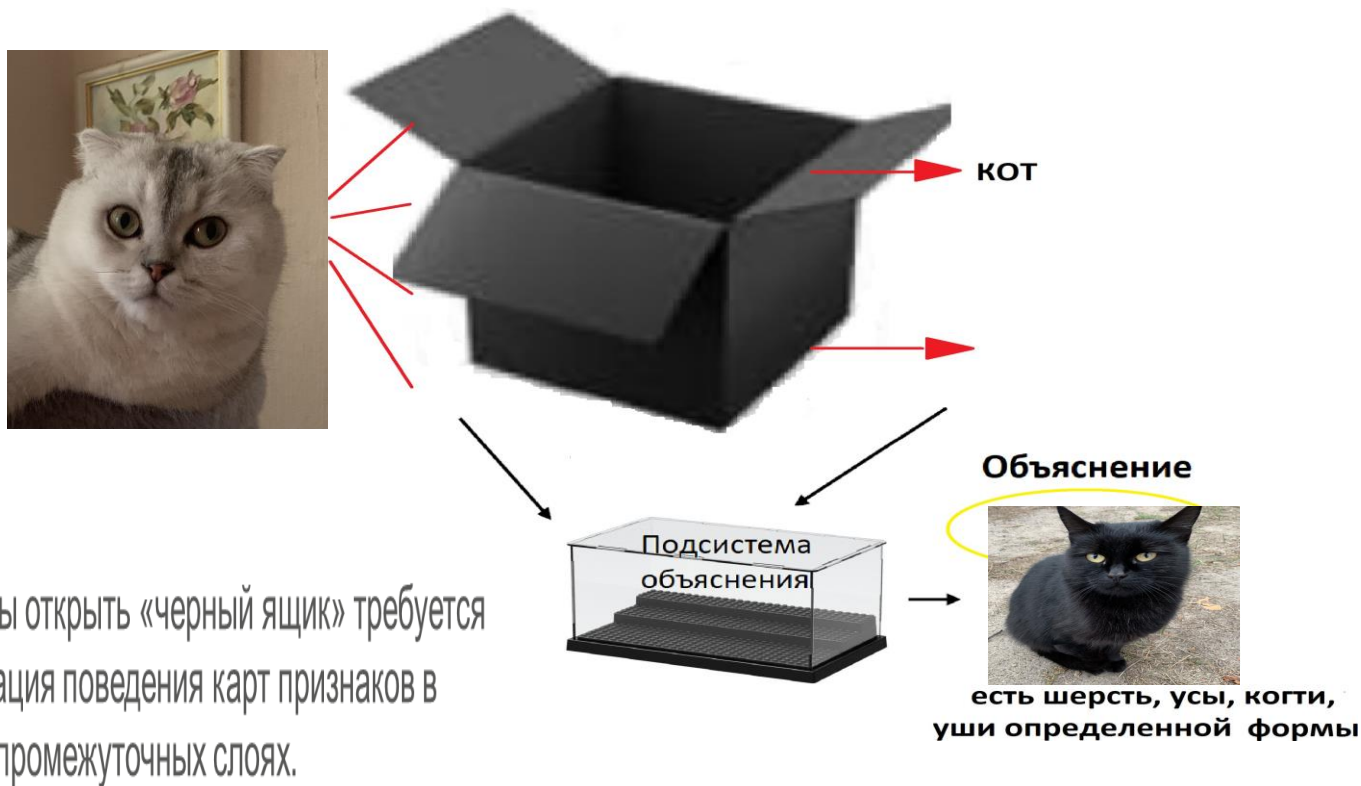
«черный ящик» **предсказывает**, что на фотографии изображен **кот** с вероятностью **0.98**



«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

Модель МО с точки зрения ПОЛЬЗОВАТЕЛЯ, НО С ДРУГИМ «ЧЕРНЫМ ЯЩИКОМ»



А НУЖНО ЛИ ОТКРЫВАТЬ «ЧЕРНЫЙ ЯЩИК» ?



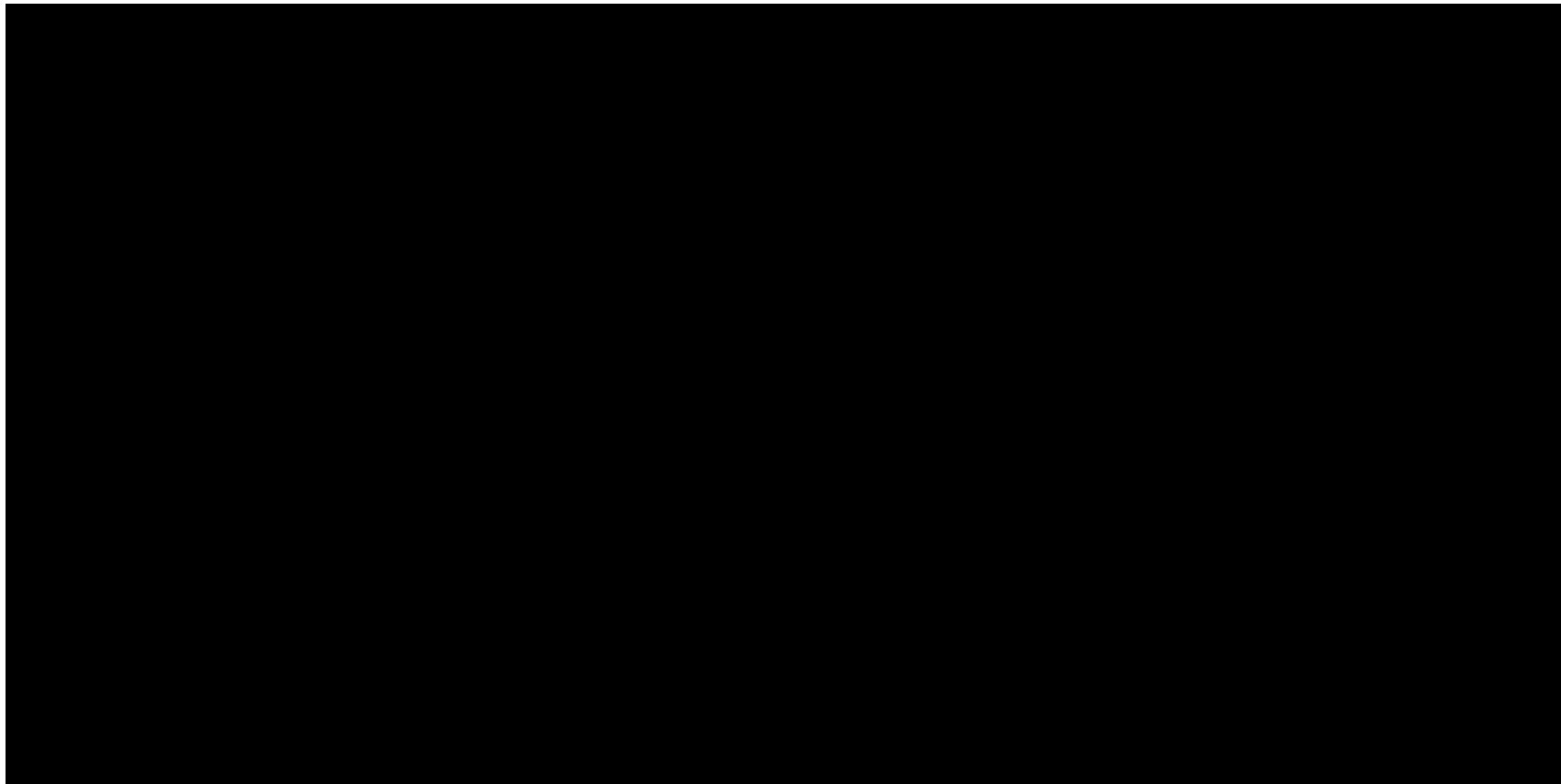
Да, нужно, так как **эффективные** (точные) модели «черного ящика» обычно сложны для понимания и не **самообъясняемые**, что снижает **уровень доверия к результатам**

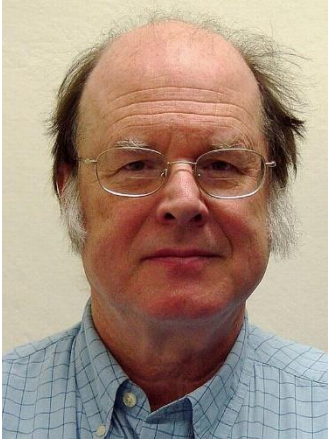
- **программирования** МТ– следование принципу хранимой программы /данных и **последовательного** выполнения множества операций булевой алгебры **ДЛЯ МОДЕЛИРОВАНИЯ НЕКОТОРОЙ ФУНКЦИИ**
- машинное обучение искусственных нейронных сетей – **«запутывание»** программы /данных

С ЦЕЛЮ «МОДКЛИРОВАНИЯ» ПРОГРАММЫ 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».





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*

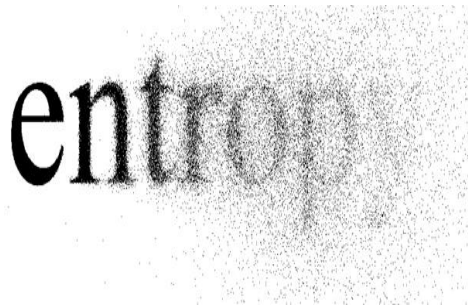
Вопрос: в чем суть «математической работы» которая производится в система ИИ с точки зрения физики?

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»

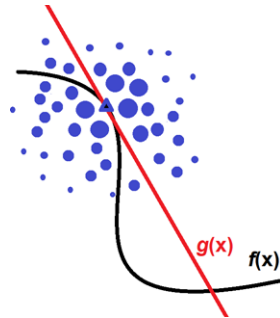
Можно ли так сформулировать *концепцию* «вычислений», так чтобы достигать «математически значимого» результата, при этом не только **потребляя энергию, но и объясняя полученные результаты ?**.

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

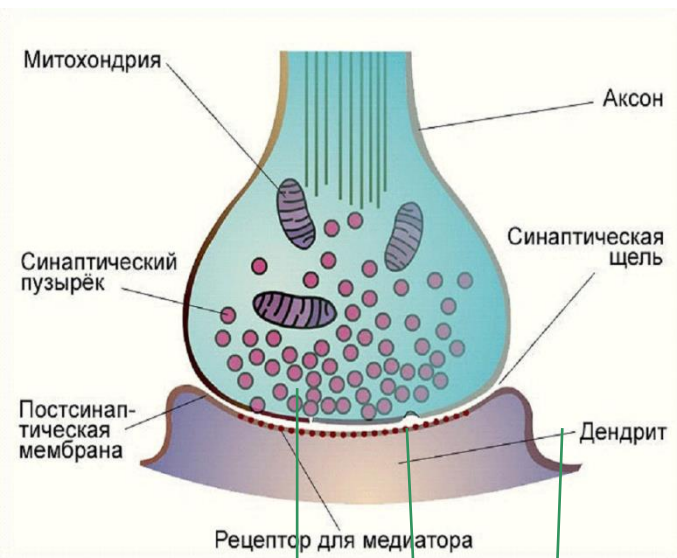


Модели
объяснения

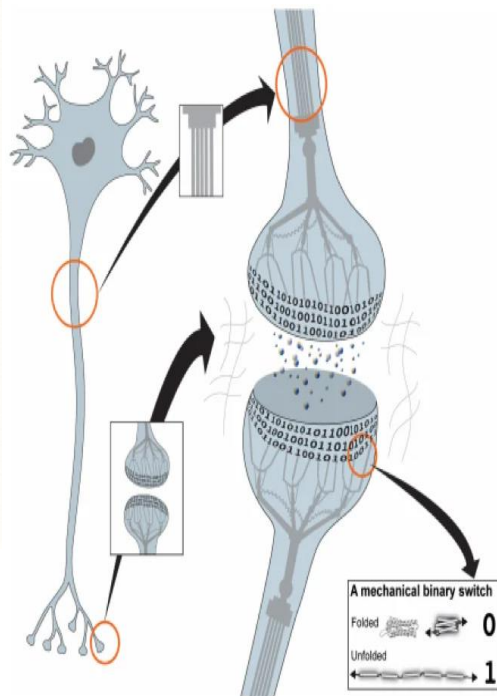
2) новый метод регуляризации с помощью использования «суррогатной» модели, используемой для объяснения того, почему полученное решение **«обратной задачи»** отвечает принятой «модели реальности»

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

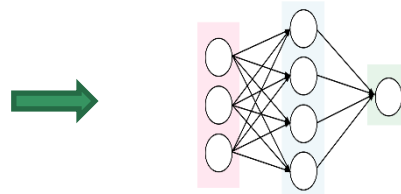
ПРИМЕР РЕШЕНИЯ : БИОИНСПИРИРОВАННАЯ СРЕДА ИНФОРМАЦИОННЫХ ПРОЦЕССОВ, ПРОТЕКАЮЩИХ В АССОЦИАТИВНОЙ ПАМЯТИ



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 «патернов» - аналога **аддитивного ситуационного базиса**



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 сгущение точек постоянной фазы, сохраняющих структуру «патерна» - образа

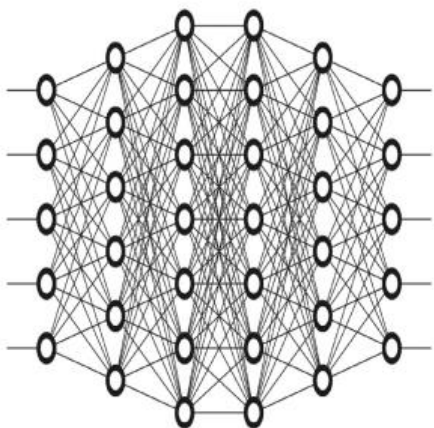
Суть научения - настройка **вычислителя** **решать** задачи исходя из требований

.... **Минимальной** диссипации энергии –
Минимального числа операций
(транзакций процессор-память)
доставляющих решение задачи

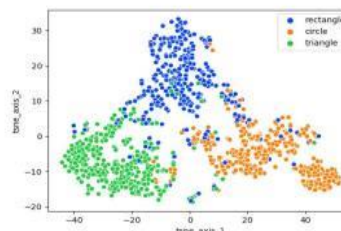
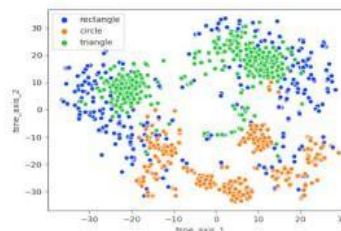
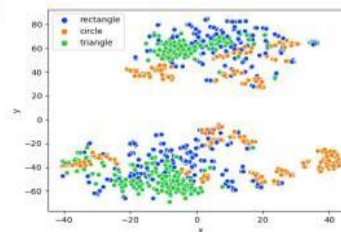
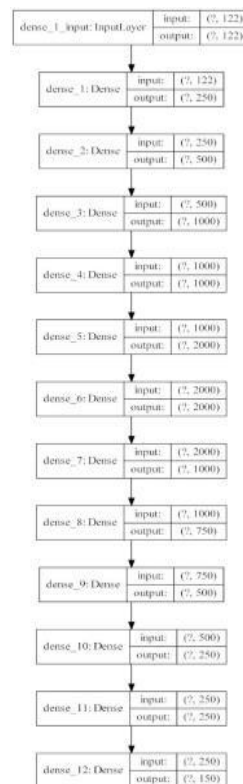
... **Объяснения** полученных результатов на
основе конечной совокупности паттернов

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



Процесс обучения ИНС приводит к
определенной структуре
аттрактора вектора состояний
слоев нейронной сети

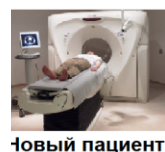
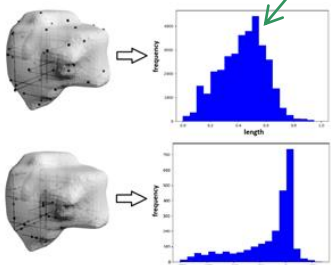


ПРИМЕР: КАК РЕАЛИЗОВАТЬ НАУЧЕНИЕ НА ПРИМЕРЕ МЕДИЦИНСКОЙ ПРОБЛЕМАТИКИ

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

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

ПРИНЦИП ПОСТРОЕНИЯ ГИСТОГРАММ НОВООБРАЗОВАНИЙ



x



$f(x)$

Злокачественное
образование

$f(x)$

x



$g(x)$

Форма - овальная
Структура - сплошная
Включения - некроз
Контур - неровный

СПАСИБО ЗА ВНИМАНИЕ!



ментально-формальный резонанс:

воображения, мышления и метафоры абстракций

идеи «обучения машин» (ОМ) :

- **Аксиома выбора** – *из непустых множеств всегда можно создать новое множество ...*
(алгоритмическая неразрешимость задач)
- **Теорема Банаха-Тарского** – *из трехмерного шара можно собрать два идентичных шара, каждый с объемом равным исходному*
(ограниченная делимость понятий и точность вычислений)
- **Теорема Геделя** – *непротиворечивая формальная система не может быть полна*
(истинность верного предположения невозможно доказать)



«ДАЛЬТНИЗМ»

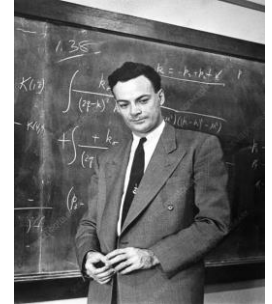
«жизненного опыта»
- решение узнаваемо,
находится быстро,
но...не точно

Парадокс «импликации»:
«из лжи может следовать
все, что угодно»



Любые обратные задачи «неустойчивы» по отношению к погрешностям в исходных данных, Поэтому нужны принципы отбора (регуляризации) «правильных решений»

А. Н. Тихонов: найти асимптотически верное решение



Р. Фейнман сказал «Заткнись и считай!»

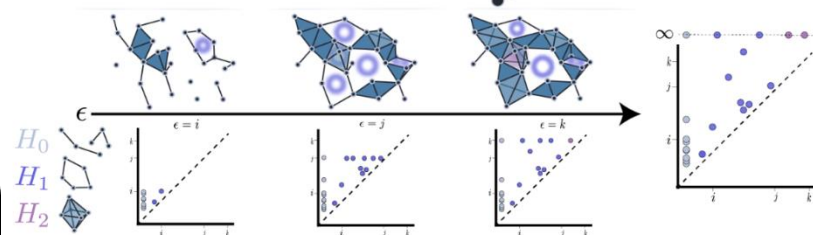
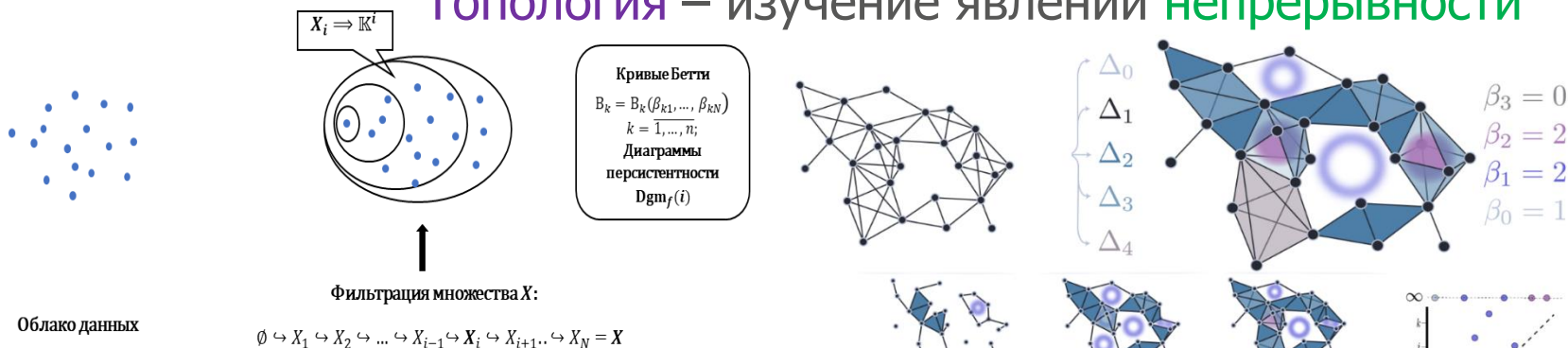
Пример из математики:

- «прямая задач: $1+2=3$ (есть одно решение)
- «обратная задача»: найти способ, как вычислить «3» (способов таких... много)

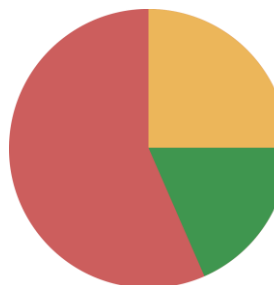
в физике проблема «обратных задач» почти решена: на «правильных» решениях должен выполняться закон сохранения энергии

- «Фундаментальный» вопрос теории ИИ: Какие «концепты» (законы) сохранения надо использовать для получения «правильных» решений задач ИИ?

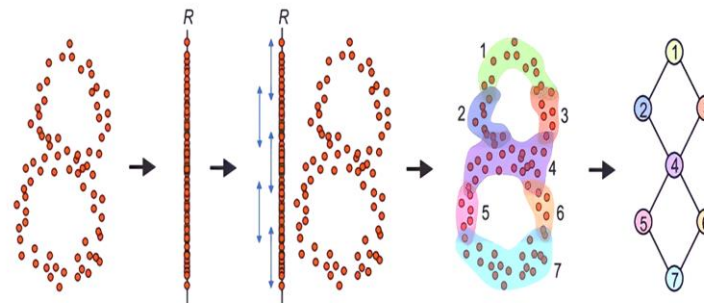
Топология – изучение явлений непрерывности



Статистика дискретных событий



Нормальное распределение (Гауссово)



гистика конечного набора данных из заданного распределения

Функции вероятности распределения событий в общем случае неизвестна

